HERIOT-WATT UNIVERSITY
DEPARTMENT OF ACCOUNTANCY,
ECONOMICS AND FINANCE

DOCTORAL THESIS

A study of the approximation and estimation of CES production functions

Elena LAGOMARSINO

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in Economics

Heriot-Watt University
Doctor of Philosophy in Economics
December 2017

Abstract

The purpose of the dissertation is to propose and explore an empirical procedure to test if a CES production function is appropriate to describe a given dataset and inform on which nested structure should be adopted when there are more than two inputs. This is particularly useful for the estimation of elasticities of substitution. The first chapter reviews the applied literature on the estimation of these elasticities and shows that Translog functions are the most popular as they are flexible enough to be adopted in various empirical applications. Conversely, Constant Elasticities of Substitution (CES) production functions are rarely employed, mostly in the computable general equilibrium (CGE) framework. Indeed, the CES production functions are based on maintained hypotheses (i.e. homogeneity, separability, and constant elasticities) which are seldom satisfied empirically. In the second chapter, we show how these assumptions can be tested, exploiting the link between the Translog and CES functions: the former can be seen as a second-order Taylor expansion of the latter. In particular, we provide the necessary and sufficient constraints on the Translog coefficients for all the feasible three-input and four-input cases. Given this information, the third chapter illustrates an empirical procedure that can be used to test whether an available dataset is consistent with a CES production technology, and, if that is the case, to determine which nested structure describes it more accurately. Finally, in the last chapter, we apply this procedure to the EU-KLEM dataset, to obtain constant elasticities of substitution for the United Kingdom.

Acknowledgements

I am very grateful to my supervisors, Prof. Mark Schaffer, Dr. Atanas Christev, and Prof. Karen Turner, for all the help, comments, and constructive critiques they have given me throughout these years. I would like to thank the external and internal examiners, Prof. Eduardo Castro and Dr. Claudia Aravena, for helpful suggestions during the viva. I would also like to thank Anna Bablyoan, Gioele Figus, Irina Myers, and Mengdi Song, fellow PhDs, for making my time here enjoyable and for providing stimulating discussions and advice; and Luca Violanti for donating his PC's spare processing power. Special thanks go to Alessandro for the unconditional practical and moral support shown to me throughout the PhD process: it is especially thanks to you that I have been able to complete this long dissertation journey.

I am also indebted to a lot of people who have commented on earlier drafts. In particular, I would like to thank Prof. Geoffrey J. D. Hewings, Prof. Kurt Kratena, Prof. Peter McGregor, Prof. Frans de Vries, and seminar participants at the Austrian Institute of Economic Research and the SGPE Residential Conferences in Crieff.

This work was produced as a postgraduate student at the School of Social Sciences of Heriot-Watt University.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Elena Lagomarsino		
School:	Social Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	PhD in Economics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:	Elena Lagomarsino	Date:	02/09/2018

Submission

Submitted By *(name in capitals)*:	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals)*:			
Method of Submission (Handed in to SSC; posted through internal/external mail):			
*E-thesis Submitted (**mandatory for final theses**)*			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.
Academic Registry/Version (1) August 2016

Declaration

This is to certify that the work contained within has been composed by me and is entirely my own work, apart from Chapter 5, which is based on a joint project with Prof. Karen Turner (I contributed 90% of this work).

This work, in the current form, has not been submitted for any other degree or professional qualification.

The main text of this thesis consists of approximately 28,633 words.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Empirical literature is often confronted with the estimation of production functions: for example, applied econometric papers regress total manufacturing output or single firms output on inputs like capital, labour, and energy to derive the value of parameters of interest (e.g. elasticities of substitution, marginal products, share parameters) that can be exploited ex post by macroeconomic and CGE models; health economics focuses on health production functions where health care, genetics, and other variables connected with lifestyle represent inputs; agricultural economics investigates the relationship between capital, labour, and land and the total output of, for instance, farming; ecological production functions link ecosystem conditions, management practices, and stressors to the production of ecosystem services; human capital papers estimate how children skills depend on parental investments and skills, and household characteristics. Nevertheless, a common denominator of this empirical work is that it neglects to provide any justification behind the choice of a particular functional form. This decision is usually based on practical needs, e.g. ease of estimation, generality of the function, convenient properties or global satisfaction of regularity conditions and superior tractability, whereas formal selection procedures are never explicitly discussed.

The purpose of the dissertation is to propose and explore an empirical procedure to test if a CES is appropriate to describe a given dataset and inform on which nested structure should be adopted when there are more than two inputs. In particular, we focus on production functions and the estimation of elasticities of substitution between inputs. This has been the objective of an impressive number of empirical papers and it is still a relevant research question, especially when the energy input is considered. For example, the decision of firms on how much to invest in energy-saving technologies is directly affected by the level at which firms can substitute away from energy and this is of utmost interest for climate policies aiming at mitigating greenhouse gas emissions.

From a review of the literature on the estimation of substitution elasticities involving two or more inputs (including energy), conducted in the first chapter of this dissertation, it emerges that this body of applied papers has been growing for almost forty years and that a general consensus on the nature of the relationship between energy and capital has yet to be reached. We also observe that, alongside the main strand of applied econometric work, the

CGE literature has recently provided estimates of constant elasticities. There are two main reasons why CGE work is increasingly interested in informing key parameters of their models using empirical data. Firstly, one of the major criticisms levelled against the CGE literature is that models are founded on key parameters in both production and consumption that lack of an empirical foundation: they are often assumed *a priori* or borrowed from previous studies. Secondly, energy and environmental CGE results have been found to be particularly sensitive to changes in the values of the elasticities of substitution between inputs of production. The main difference between the two strands of literature is the functional form they employ to describe the input-output relationship: while the first is based on Translog cost functions for its flexibility and the ease with which its share equations and Allen elasticities can be derived and estimated, CGE literature favours Constant Elasticity of Substitution (CES) functions for their convenient characteristics and global validity. We conclude that applied research should pay particular attention to the assumptions they make about model specification, the type of elasticity they choose, and the econometric technique they apply. Moreover, we warn researchers to be careful in the use of CES production functions as from an empirical standpoint these functional forms are very restrictive: they are based on strong maintained hypotheses on technology and inputs (i.e. homogeneity and strong separability, constant elasticities) which have often been rejected in real data applications.

This calls for an empirical procedure to test if a nested CES is appropriate to describe a given dataset and which nested structure is the most realistic. A potential idea, investigated in this dissertation, is to base the procedure on a flexible functional form on which the CES maintained hypotheses could be tested. The most suitable candidate is the Translog as the connection with the CES is straightforward: when the Translog coefficients satisfy the CES hypotheses, it can be interpreted as a second order Taylor approximation to an arbitrary CES.

In the second chapter, we look at which constraints should be imposed on a Translog production function to test for separability. Although a general indication on how to derive input separability conditions for a Translog function can be found in Berndt and Christensen (1973b), only simple separability structures and a limited number of inputs have been considered so far. We outline a simple method that can be used with any n-input Translog functions to identify the number and the form of the necessary and sufficient restrictions required to test for various forms of input separability. This is based on the comparison between the Translog and the nested CES by means of a linear approximation of the latter: the way inputs are nested in a CES reflects a specific input separability structure. In the show for the first time how to resolve the multivariate second-order Taylor expansion of nested CES functions. Furthermore, we explicitly provide separability constraints for all the feasible three-input and four-input Translog cases.

The description of a potential empirical procedure is discussed in the third chapter, which tries to answer the following question: on which basis should a researcher opt for a CES? A Monte Carlo simulation environment is exploited to assess how often the various phases of the procedure correctly recognize the functional form of the production function assumed in the data generating process. The first phase consists in a number of inference tests performed on the Translog coefficients in order to understand if the Translog is homogeneous and separable, i.e. if some of the CES assumptions are satisfied. A failure to reject the tested restrictions represents a first indication that a CES could be the appropriate function to describe the input-output relationship. Moreover, with more than two inputs, the test also informs on which nested CES structure more closely approximates the true one. In empirical applications, the results of this phase can deliver two outcomes. On the one hand, the results may indicate that we fail to reject some of the maintained characteristics of the non-linear CES and, thus, the procedure concludes that a CES is not appropriate for that specific dataset. On the other hand, results may be in favour of a CES model, and a second phase should be used to understand if the underlying model is a non-linear CES or just its approximation. The second phase consists in both a graphical analysis and formal tests based on selection criteria. Once the Translog is estimated, it is possible to derive its point substitution elasticities and prediction intervals around them. If we observe peaks in their distribution around a small range of values and narrow prediction intervals, we can conclude that the dataset supports the hypothesis of a constant elasticity (i.e. a CES structure is appropriate). Formal tests consist in computing different selection criteria to determine which of the two rival models performs better.

Finally, in the fourth chapter, we apply the proposed procedure to the EU-KLEM dataset, to understand whether a nested CES production function is adequate and to obtain an indication on which nested structure is the most appropriate for the data considered. Given the finite number of panels and the long time-series component, stationarity, cointegration, and contemporaneous correlation are accounted for. Findings from this first attempt of applying the procedure suggest that a nested CES where energy and capital inputs form an inner nest that is combined with labour and materials at an outer level of production is the one that more closely describes the UK production technology with inner and outer elasticities of 0.88 and 0.47, respectively.

Chapter 2

Estimating Elasticities of Substitution between Energy and Other Inputs

A review of the literature

2.1 Introduction

The first studies on the substitutability between production inputs date back to the 1930s when Hicks (1932) and Robinson (1933) formalized two independent concepts of elasticity of substitution between capital and labour. Only in the 1970s, energy was recognised as a key input in production. Indeed, after the burst of the oil crisis, and the subsequent embargo in 1973, the price of oil quadrupled, and this prioritized the analysis on the level at which energy could be substituted with other factor inputs. Of particular interest was the relationship between capital and energy: if the two inputs were complements, an increase in energy prices would have led to a downturn in capital formation and, hence, to a slowdown of the economic growth. On the contrary, if the two inputs were substitutes, a rapid formation of capital would have balanced the limited use of energy resources and helped to avoid a recession. Since then, and following Hudson and Jorgenson (1974) and Berndt and Wood (1975), a vast body of literature has been trying to provide empirical estimates of the level of substitution between factor inputs.

Nowadays this research question is still very relevant. Firstly, scarcity of non-renewable resources may lead to a sharp increase in their relative price and, hence, to the same type of concerns raised during the oil crisis. Secondly, the decision on how much to invest in energy-saving technology is driven by the level at which one can substitute away from energy and this is of utmost interest for the climate policies aiming at mitigating greenhouse gas emissions.

From a comprehensive analysis of the literature what emerges is that, although improvements have been made both in the estimation procedures and in the availability of appropriate dataset, the research outputs are very discordant. The most intense debate concerns the nature of the relationship between capital and energy. Previous reviews tried to individuate the reasons for so many diverging results. Apostolakis (1990) asserted that the problem lies in the different kind of data used in the papers: while time-series studies show the short-term relationship between capital and energy, cross-section analysis capture the long-term one. On the contrary, Thompson and Taylor (1995) sustained that the disparity of results is due to the different substitution elasticities considered and offered a proof showing that, once previous results are translated into Morishima elasticities, they all support energy/capital substitution. Finally, the meta-analysis study of Koetse et al. (2008) justified the heterogeneity in the results with the different model specifications, data characteristics and economic context.

In light of the conclusions of previous reviews, in this chapter we examine the five aspects that help explaining the existence of diverging results: the assumptions regarding the production function, the type of elasticities of substitution, the data characteristics, the econometric methods, and the economic context. For each of them we describe the choices that have been made so far in the existing literature and the consequences they had on results. The purpose is to provide a guidance for future research illustrating the options available and the most recent advancements.

Furthermore, for the first time, we include in the review a group of applied works linked to Computable General Equilibrium (CGE) literature. Indeed, in recent years, CGE modellers have developed an interest in the estimation of substitution elasticities for two main reasons: they have been long criticized for the lack of empirical foundation that characterizes these parameters and they found them to play a decisive role on simulation results in the energy/environmental context.[1]

The main difference between the typical applied econometric papers and those intended for CGE applications is the functional form they adopt to describe the production technology: while the first resort to flexible functional forms (FFF) for their general and convenient applicability, the second employ nested constant elasticity of substitution (CES) functions for their global validity and greater tractability. Results are difficult to compare in terms of magnitude as FFF are characterized by non-constant elasticities but they can be regarded to shed light on whether they predict the relationship between inputs to be of complementarity or substitutability.

Obviously, the type of choices available for each of the aforementioned aspects could be different when using a nested CES production function (i.e. assumptions, elasticities of substitution, and econometric method), thus for each of them we will consider separately the two functional forms.

[1] See for example Lecca et al. (2011) for an analysis regarding energy rebound effects.

As summarized in Table 2.1, this literature review covers on forty works. The chapter begins with a brief revision of the two class of functional forms to highlight the main differences in Section 2.2. Then each section is dedicated to one of the aspects: in Section 2.3 we look at the assumptions, Section 2.4 at the elasticities of substitution, Section 2.5 at the level of data collection, Section 2.6 at the estimation techniques, and Section 2.7 at the economic context. Finally, Section 2.8 concludes.

2.2 Functional forms

Production functions describe the technology that transforms factor inputs into output. Formally:

$$Q = f(x_1, \dots, x_n, T) \tag{2.1}$$

where Q represents final output, x_i with $i, j = 1, ..., n$ represent production inputs and T is a time variable ($T = 1, 2, \dots$) used to analyse technical progress.

There are two categories of functional forms that can be employed to describe a production function with three or more inputs: the FFF and the CES functions. While the first was used by nearly all studies written between 1975 and 2008,[2] the second characterizes most of the recent ones. The reason originates in the fact that most of the recent papers aim at providing estimates to incorporate in CGE models, which are traditionally based on production functions described by nested CES .

The two categories of functional forms differ in various ways. When assessing which of them is the most appropriate, one should bear in mind the trade-off between generality and global validity: while FFFs allow more flexibility and generality, they are not guaranteed to be well-behaved in all production domain; on the contrary, nested CES always satisfy production regularity conditions but are based on a series of maintained assumptions on technology, inputs, and substitution elasticities that are not always realistic. This issue will be further developed in this section together with a brief introduction to the two most used functional forms, the Transcendental logarithmic (i.e. Translog) and the CES.

2.2.1 Transcendental Logarithmic

The FFFs category was introduced after Diewert's (1971) definition of flexibility and includes more than fourteen different functional forms. The most frequently used are the

[2]With the exceptions of Prywes (1986), Chang (1994) and Kemfert (1998).

Transcendental Logarithmic (hereafter Translog), due to Christensen et al. (1973), the Generalized Leontief, due to Diewert (1971) and the Generalized Cobb-Douglas, due to Diewert (1973). As all but three[3] (see Table 2.1, column Fcn) of the thirty-one selected studies that use FFFs employ a Translog function, we focus on this particular functional form.

A Translog production function can be written formally as:

$$\ln(Q) = \ln(a_0) + \sum_{i=1}^{n} a_i \ln(x_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \ln(x_i) \ln(x_j) \tag{2.2}$$

where a_0 is the efficiency parameter and a_i and a_{ij} are unknown parameters to be estimated. When $a_{ij} = 0$ for all i and j, the Translog production function reduces to a Cobb-Douglas.

Translog production functions have at least three interpretations: they can be seen as exact representations of the true production technology, as second order Taylor approximation to a CES function, or as second order Taylor approximation to an unknown underlying production function. When the Translog is seen as a linearisation, it is important to remember that a remainder (i.e. an approximation error) should be attached to equation (2.2) and that its magnitude increases as we move away from the approximation point.

The Shephard duality theorem[4] allows the researcher to employ a cost function that corresponds to the production function and reflects the same production technology. The decision to exploit a production or a cost function has significant repercussions on final results. Indeed, Burgess (1975) underlined how Translog functions are not self-dual and showed how this implies that the elasticities resulting from the estimation of a production or a cost functions may differ substantially. Traditionally, applied studies on the estimation of substitution elasticities have favoured cost functions for two main reasons. First, the independent variables in the estimation are prices whose exogeneity is more justifiable than for quantities. Second, it is possible to base the estimation procedure on a system of cost share functions, rather than on the cost function itself which, like its dual production function, could be subject to multicollinearity. However, these two reasons can be called into question. Whereas prices could perhaps be considered exogenous for a single firm

[3]Danny et al. (1978) and Ilmakunnas and Torma (1989) employ a Generalized Leontief production function and Magnus (1979) an Extended Generalized Cobb-Douglas production function. In particular, Magnus (1979) in his study on the Netherlands manufacturing sector for the period 1950-76, compared the estimation results derived by means of a Generalized Cobb-Douglas and a Translog cost function and concluded that the estimates are comparable both in terms of sign and magnitude.

[4]As recalled by Diewert (1971, p.482): *"The Shephard duality theorem (1953) states that technology may be equivalently represented by a production function, satisfying certain regularity conditions, or a cost function, satisfying certain regularity condition"*.

or a group of firms, it may be unrealistic to assume that a whole country industrial or manufacturing sector does not influence selling prices, or that the output price has no effect on the price of capital and material inputs.[5] Moreover, opting for a cost function implies imposing the assumptions of perfect competitive markets and price homogeneity which cannot always be guaranteed. Finally, deriving marginal productivities and the relative standard errors becomes very problematic when using cost shares. Nevertheless, all but one[6] of the surveyed papers based on the Translog have opted for a cost function.

For a four input model, a twice differentiable Translog cost function can be written as:

$$\ln C = \ln(b_0) + b_q \ln(Q) + b_T \ln(T) + \sum_{i=1}^{n} b_i \ln(P_i) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} \ln(P_i)\ln(P_j) + \\ + b_q q(ln(Q))^2 + b_T T(ln(T))^2 + \sum_{i=1}^{n} b_{iq} \ln(P_i)\ln(q) + \sum_{i=1}^{n} b_{iT} \ln(P_i)\ln(T) \quad (2.3)$$

where C is total cost, P_i is the price of input i, $i = 1, ..., n$, and T is a time variable representing technical progress.

According to neoclassical theory, cost functions must be homogeneous of degree one in input prices, fulfil the symmetry requirements and be non-decreasing and concave in input prices. Homogeneity in prices and symmetry[7] means that the following constraints need to be satisfied:

$$\sum_{i=1}^{n} b_i = 1, \quad \sum_{i=1}^{n} b_{ij} = \sum_{j=1}^{n} b_{ij} = 0, \quad \sum_{i=1}^{n} b_{iq} = 0, \quad \sum_{i=1}^{n} b_{iT} = 0, \quad b_{ij} = b_{ji}. \quad (2.4)$$

Cost functions, like production functions, are not globally well-behaved. Thus, regularity conditions need to be imposed and tested: there needs to be a region in the input space large enough to guarantee that the production function is appropriately represented. Monotonicity and concavity should be tested at each observation after the estimation. Positive fitted input share equations indicate that the cost function is monotonic in prices. To check for concavity, the bordered Hessian matrix obtained from the coefficients estimation must be found to be negative semi-definite.[8] Although many studies on the estimation of elasticities substitution with a Translog function assumed well-behaved production functions without

[5]The only paper treating prices as endogenous is the one by Berndt and Wood (1975) who constructed instruments and estimated a 3-stage least squares to deal with the issue.

[6]Norsworthy and Malmquist (1983).

[7]According to Young theorem on the equivalence of second class partial derivatives.

[8]Christev and Featherstone (2009) showed that in Translog cost functions the curvature of the Hessian matrix can be checked also through the matrix of the Allen elasticities of substitution.

testing for it (Ozatalay et al., 1979, Norsworthy and Malmquist, 1983, Moghimzadeh and Kymn, 1986, Garofalo and Malhotra, 1988, Hisnanick and Kyer, 1995, Christopoulos, 2000, Khiabani and Hasani, 2010, Kim and Heo, 2013), others have verified if their estimated Translog satisfied the regularity conditions. Among these, few found they were satisfied on all the domain (Berndt and Wood, 1975, Griffin and Gregory, 1976, Fuss, 1977, Turnovsky et al., 1982, Burki and Khan, 2004, Roy et al., 2006) but in numerous other cases monotonicity or the curvature conditions were rejected for at least some of the observations in the dataset. The consequent responses have been manifold: exclude all the observations where the monotonicity condition were not satisfied but keep those where isoquants convexity was rejected (Medina and Vega-Cervera, 2001), remove the sectors/countries that were more affected by the rejection (Field and Grebenstein, 1980, Medina and Vega-Cervera, 2001), proceed with the estimation ignoring the rejection (Dargay, 1983, Hesse and Tarkka, 1986, Nguyen and Streitwieser, 1999).

The derivation of conditional input demand functions, and consequently cost shares, is very straightforward. If we assume neoclassical markets, we can use Shephard's Lemma (i.e. the partial derivative of the cost function with respect to input prices yields the optimal level of inputs as a function of prices and output): for all $i, j = 1, ..., n$,

$$\frac{\delta C}{\delta P_i} = x_i^* \tag{2.5}$$

where x_i^* is the conditional demand of input i. We can logarithmically differentiate the total cost function with respect to input prices,

$$\frac{\delta lnC}{\delta lnP_i} = \frac{\delta C}{\delta P_i}\frac{P_i}{C} = b_i + \sum_{i=1}^{n} b_{ij}\ln(P_j) + b_{iq}\ln(q) + b_{iT}\ln(T) \tag{2.6}$$

and substitute (2.5) in (2.6) to obtain the input cost shares

$$S_i = x_i^*\frac{P_i}{C} = b_i + \sum_{i=1}^{n} b_{ij}\ln(P_j) + b_{iq}\ln(q) + b_{iT}\ln(T) \tag{2.7}$$

where C is total cost, i.e. $C = \sum_{i=1}^{n} P_i x_i$, and output price is normalised to 1.

Following Uzawa (1962), the Allen Partial Elasticities of Substitution (AES)[9] are given by:

$$\sigma_{ii} = \frac{b_{ii} + S_i^2 - S_i}{S_i^2}, \qquad i = 1, ..., n \tag{2.8}$$

$$\sigma_{ij} = \frac{b_{ij} + S_i S_j}{S_i S_j}, \qquad i, j = 1, ..., n \qquad i \neq j. \tag{2.9}$$

Despite being not globally well-behaved and, depending on the interpretation, subject to an approximation error, Translog functions have at least two appealing characteristics. Firstly, they are log-linear in their inputs and this is particularly convenient for their econometric estimation. Secondly, they have neither built-in assumption on inputs and technology (e.g. homogeneity, homotheticity, separability) nor an assumption of constancy of the elasticities of substitution: as a result, they are general and flexible enough to be adaptable to any type of dataset.

2.2.2 CES functions

The original two-input CES production function was introduced by the Stanford group around Arrow et al. (1961). Subsequently, several attempts to an n-input generalization where proposed, but only two stood out: the one-level n-input CES by Blackorby and Russell (1989) and the nested CES by Sato (1967). The first one is a very straightforward extension of the 2-input case where all inputs are combined at the same level of production and share the same degree of substitutability. The nested CES is a multi-factor function where n-inputs are nested at different levels of production according to a pre-determined structure and eventually combined to form final output. The nested structures range from the case of a 2-level CES with a single inner nest to a n-level CES where pairs or groups of inputs are nested at different levels. Nested CES allow a greater degree of adaptability as different combinations of inputs can have different degrees of substitutability between them. For this reason, they have been chosen by all the studies considered in this survey that are not using a FFF (i.e. in CGE modelling).

A nested CES of the form $((x_1, x_2), x_3)$ is formulated as:

$$Q = \lambda\Big(\delta X^{-\rho} + (1-\delta)x_3^{-\rho}\Big)^{-\frac{\nu}{\rho}} \tag{2.10}$$

with

$$X = \lambda_x\Big(\delta_x x_1^{-\rho_x} + (1-\delta_x)x_2^{-\rho_x}\Big)^{-\frac{1}{\rho_x}} \tag{2.11}$$

[9]For this reason sometimes called Allen-Uzawa elasticities.

where $\lambda \in [0, +\infty)$ and $\lambda_x \in [0, +\infty)$ are the efficiency parameters, $\delta \in (0, 1)$ and $\delta_x \in (0, 1)$ are share parameters, $\nu \in (0, +\infty)$ is the scale parameter and $\rho \in (-1, +\infty)$ and $\rho_x \in (-1, +\infty)$ are substitution parameters. The elasticity of substitution between input x_3 and the composite input $\mathbf{X}$ is given by $\sigma = 1/(1 + \rho)$ and the elasticity of substitution between input x_1 and input x_2 is given by $\sigma_x = 1/(1 + \rho_x)$.

A nested structure can be selected *a priori* or data driven. In the early studies of Prywes (1986) on two-digit US manufacturing industries and Chang (1994) on the Taiwan manufacturing sector, a three-level CES is assumed *a priori* as the function that best represented the data. The selected structure for the three-level CES was of the form $(((K, E), L), M)$, i.e. the first nest is composed by capital and energy, the second nest is a combination of the first nest and the labour input and the outer nest combines the second nest with the materials input.

The development of general equilibrium models for climate modelling brought about changes in the way nested CES production functions were specified. In 1998, Kemfert, in a study based on the West Germany manufacturing sector between the years 1960-93, showed how to empirically choose which structure specification is the one that best fits the data. She estimated three alternative nested structures[10] for a two-level CES production function and looked a the R^2 statistic in each case: the nested structure with the largest R^2 was then selected as the one best fitting the data.

The subsequent studies,[11] apart from Koesler and Schymura (2015),[12] replicated Kemfert's (1998) R^2 approach.[13] However, according to Baccianti (2013), the use of the R^2 statistic could be inappropriate because, as it will be showed in Section 2.6, when using a minimization approach, the alternative R^2 statistics are based on econometric models using different dependent variables. We also argue that the use of this selection criteria implies that the researcher believes the true unknown functional form to be consistent with a CES which might not be the case given that CES is highly restrictive.

Zha and Zhou's (2014) study on the steel sector in China in the period 1995-2008 uses a new approach for determining the nested structure. They estimate a 3-input Translog function and compute the elasticities of substitution between them. To build a nested CES, they assume that the inner nest is composed by the two inputs with the largest elasticity on average. Such an assumption though is not theoretically justified.

[10] $((x_1, x_2), x_3)$, $((x_1, x_3), x_2)$, $((x_2, x_3), x_1)$.

[11] van der Werf (2008), Okagawa and Ban (2008), Ha et al. (2012), Baccianti (2013).

[12] The authors, for a panel data of 40 countries in the period 1995-2006, estimated the structure $(((K, L), E), M)$.

[13] Baccianti (2013) considers also a one level nested structure (K, L, E)

Nested CES functions have the advantage of being globally well-behaved by construction. Nevertheless, they have some drawbacks. Firstly, they imply that the data under analysis satisfy a number of maintained assumptions on inputs (i.e. homogeneity, homotheticity and separability) and that the underlying elasticities of substitution between inputs are constant. Secondly, they are non-linear which makes their estimation not straightforward. Thirdly, they implicate an *a priori* decision on which nested structure to adopt.

We are aware of only one empirical work in which the two functional forms are compared, namely Chang (1994). The author's conclusion is that the Translog cost function and the CES production function for the Taiwanese manufacturing sector in the period 1956-71 produce AES that are not significantly different in magnitude.

2.3 Assumptions

2.3.1 Homogeneity, homotheticity and weak separability

It is a common theme in economics that the fewer assumptions are made, the more general a model becomes. Even though the Translog cost function only requires symmetry and linear homogeneity in prices, sometimes more assumptions are specified to simplify computations. In order to satisfy homotheticity,[14] the Translog cost function defined in (2.3) must satisfy the following additional restrictions on its estimated coefficients:

$$b_{iq} = 0 \qquad \forall\ i. \tag{2.12}$$

For homogeneity in prices and output, the following additional constraint is required:

$$b_{qq} = 0 \qquad \forall\ i. \tag{2.13}$$

Homogeneity is therefore a special case of homotheticity: if a function is homogeneous it is also homothetic. Homogeneity implies that average costs are constant. Linear homogeneity, or constant returns to scale of the dual production function, is attained when:

$$b_q = 1. \tag{2.14}$$

[14]Homotheticity implies that the cost function is separable in output and factor prices. If the cost function is homothetic, input demand functions do not depend on the output level.

In many cases, separability[15] is also assumed. The main reason in that, frequently, the production factors need to be limited to capital, labour, and energy because of the unavailability of the material input data. Another reason is the disaggregation of inputs: some papers explored the substitution degree between different types of energy or capital inputs. For a three inputs Translog cost function, separability of the form $((x_1, x_2), x_3)$ is guaranteed by the following constraints:[16]

$$b_1 b_{23} - b_2 b_{13} = 0 \qquad \text{and} \qquad b_{1m} b_{23} - b_{2m} b_{13} = 0 \qquad \text{with} \qquad m = 1...n. \qquad (2.15)$$

Half of the studies which adopted a Translog function assumed homotheticity (see Table 2.1, column CRTS), and most of the remaining found empirical evidence in favour of a non-homothetic cost function. Dargay (1983) compared the estimation results from both the homothetic and non-homothetic versions of the Translog cost function and found the latter to provide estimates which are smaller in magnitude.

As the test for constant returns to scale implies the estimation of the Translog cost function itself, only three studies proceeded with it, namely Iqbal (1986), Khiabani and Hasani (2010) and Haller and Hyland (2014). Whereas more than half of the surveyed papers assumed them, the three studies find that constant returns were not supported by the data they analysed.

Regarding separability, all the studies which disregarded the material input had to assume it to be separable from the remaining inputs. Eleven studies disaggregated one or more of the inputs (see Table 2.1, column Disag), e.g. energy disaggregation in electricity and fuel, and only four of them (Hazilla and Kopp, 1986, Moghimzadeh and Kymn, 1986, Garofalo and Malhotra, 1988, Hisnanick and Kyer, 1995) performed tests to verify if the sub-inputs are separate from the others and can be combined in an intermediate input. Only Hisnanick and Kyer (1995) did not reject separability and showed how the elasticities of disaggregated inputs are comparable to each other. Furthermore, of these eleven studies, all but Fuss (1977) and Pindyck (1979) estimated the disaggregated model. The two mentioned authors, instead, built an intermediate input using a weakly homothetic separable function on the basis of which they successively generated the aggregated model. A few other papers tested for strong separability conditions. Berndt and Wood (1975) and Chung (1987),

[15]Consider a twice differentiable strictly concave homothetic production function $f(x) = f(x_1, .., x_n)$ whose input are partitioned into R mutually exclusive subset $[N_1, .., N_R]$. $f(x)$ is separable with respect to a partition R if the marginal rate of substitution between x_i and x_j for any subset N_s with $s = 1, \ldots, R$ is independent of the quantities outside N_s (Christensen et al., 1973). Separability implies equality of the Allen elasticities of substitution: $\sigma_{ik} = \sigma_{jk} \quad (i, j \in N_s, k \notin N_s)$.

[16]See Berndt and Christensen (1973b).

looking at the same dataset on the US manufacturing sector for the period 1947-71 and using two different approaches, failed to reject the separability conditions only for the case $((KE)(LM))$. Medina and Vega-Cervera's (2001) results indicated that a value added aggregate can be separated from other inputs for the three countries she analysed and Griffin and Gregory (1976) obtained the same results for both their US and European models. On the contrary, Roy et al. (2006) found no evidence in favour of value added separability in his study based on developing countries data.

Nested CES functions, as mentioned above, are very restrictive as they are built on the maintained hypotheses of homotheticity, homogeneity, and separability. An additional assumption that could be made on when using this functional form is linear homogeneity imposing the scale parameter v to equal one.

2.3.2 Technical change

As recalled by Binswanger (1974, p.964), the Hicks definition of technical change is the following: *"Technical change is said to be neutral, labour-saving, or labour-using depending on whether, at a constant capital-labour ratio, the marginal rate of substitution stays constant, increases or decreases. Mathematically this can be stated as follows:*

$$\frac{d}{dt}MRS = \frac{d}{dt}\frac{f_K}{f_L} = -\frac{d}{dt}\frac{K}{L} \tag{2.16}$$

where f_K and f_L stand for the marginal products and the capital-labour ratio is held constant".

A distinction between those studies which employed a FFF and those which adopted a CES formulation is required also at this stage. As illustrated by Binswanger (1974), non-neutrality implies that there is, over time, a change in the cost shares of inputs. When the change in b_{it} is positive (negative), the technical change is said to be factor i-using (i-saving). There are only seven papers which tested for technical bias and did not assume Hicks-neutrality (see Table 2.1, column HN), and all of them significantly rejected it. Apart from Burki and Khan (2004) and Roy et al. (2006), who estimated both the cost function and the cost shares, in all other studies the effect of technological change has been evaluated at the optimal factor use. Among them, the two studies on developing countries (Burki and Khan, 2004, Khiabani and Hasani, 2010) find evidence of energy-using technologies together with, respectively, raw-materials, and capital saving technologies. Also Hesse and Tarkka (1986), in their study on eight European countries over the period 1960-80, found technical change to be energy-using (and labour-saving) and they also showed that, relaxing the assumption of Hicks-neutrality, capital and labour from substitutes become complements. Roy et al. (2006) obtained insignificant estimates for the technological coefficients. Danny et al. (1978) reckoned that Hicks-neutral elasticities estimates are upward biased but that the sign of the relationships is invariant to technological change.

In the nested CES framework, the early studies of Prywes (1986) and Chang (1994) introduced a parameter measuring technical change. In particular, Prywes (1986) added a total factor productivity term at each nest so that the three sub-functions were Hicks-neutral but the overall nested CES production function could be affected by technical changes.

The first study which introduced factor-augmenting technical change in a nested CES production function was published by van der Werf (2008), who added the terms A_i to the traditional specification to represent the factor specific levels of technology in his panel estimation on twelve countries for the period 1978-96. Formally:

$$q = \lambda\Big(\delta\mathbf{Z}^{-\rho} + (1-\delta)(A_{x3}x_3)^{-\rho}\Big)^{-\frac{\nu}{\rho}} \tag{2.17}$$

with

$$Z = \Big(\delta_x(A_{x1}x_1)^{-\rho_x} + (1-\delta)(A_{x2}x_2)^{-\rho_x}\Big)^{-\frac{1}{\rho_x}} \tag{2.18}$$

In his paper, van der Werf (2008) tested also whether a total factor productivity representation of technology was more appropriate than input specific technological trends (i.e. $A_{x1} = A_{x2} = A_{x3}$). His results indicated that a model with input-neutral technological change is rejected by the data. The subsequent literature[17] adopted the same way of modelling technical change.

2.4 Elasticities of substitution

The concept of elasticity of substitution (ES) was introduced by Hicks (1932) and applied to the two inputs capital and labour. Formally, it is measured as the ratio of two inputs with respect to the ratio of their marginal product and provides *"a measure of the ease with which the varying factor can be substituted for others"* holding output constant (Hicks, 1932, p.117). When the quantities of inputs are optimal, the ES can be written as:

$$\sigma = \frac{\partial \ln \frac{K}{L}}{\partial \ln \frac{P_K}{P_L}} \tag{2.19}$$

[17] Okagawa and Ban (2008), Ha et al. (2012) and Baccianti (2013).

The closer the ES is to zero, the less the two inputs can be substituted.[18]

Allen (1934) generalized this concept to a multi-factor production function in two separate ways. The first way led to the development of the so-called Hicks elasticity of substitution (HES)[19] which can be computed by applying the original elasticity of substitution to each pair of inputs holding the quantities of the others and output constant. These are considered short-run elasticities because other inputs are not allowed to adjust. HES can be written as:

$$\sigma_{ij}^{HES} = \frac{dln(\frac{x_i}{x_j})}{dln\frac{f_j}{f_i}} = \frac{d\frac{x_i}{x_j}\frac{f_j}{f_i}}{d\frac{f_j}{f_i}\frac{x_i}{x_j}}, \tag{2.20}$$

where $f(x_1, x_2)$ is the production function that it is assumed to produce output q, and f_i and f_j are the partial derivatives of the production function with respect to input i and j respectively, i.e. the marginal products of the two inputs. Using total differentiation and Young's Theorem (i.e. $f_{12} = f_{21}$) we find that HES can also be expressed as:

$$\sigma_{ij}^{HES} = \frac{(f_i x_i + f_j x_j)}{x_i x_j} \frac{f_1 f_2}{(2f_{12}f_1 f_2 - f_{22}f_1^2 - f_{11}f_2^2)} \tag{2.21}$$

The second way led to the introduction of the partial elasticities of substitution, which have been successively re-investigated by Allen (1938) and by Uzawa (1962). Formally, AES are defined as:

$$\sigma_{ij}^{AES} = \frac{\sum_{k=1}^{n} f_k x_k}{x_i x_j} \frac{|D_{ij}|}{|D|} \tag{2.22}$$

where $|D|$ is the determinant of the bordered Hessian matrix D formed by the production function and $|D_{ij}|$ represents the cofactor of the ikth term in the Hessian matrix. A negative AES indicates complementarity and a positive AES indicates substitutability. The main difference between HES and AES is then given by the fact that AES holds only output constant.

An interesting alternative expression for the AES based on cost functions is:

$$\sigma_{ij} = \frac{C(p,q)C_{ij}(p,q)}{C_i(p,q)C_j(p,q)} \tag{2.23}$$

where $C(p, q)$ is the cost function, p is a vector of n inputs prices ($i, j = 1, \ldots, n$) and the subscript on the cost function refers to the partial derivative with respect to that particular input price.

[18] $\sigma = 0$ is the case of perfect complementarity. $\sigma = 1$ in the case of a Cobb-Douglas production function.
[19] Also known as direct elasticities of substitution.

Allen (1938) showed that the cross-price elasticities of demand E_{ij} can be written in terms of AES as:

$$E_{ij} = S_j \sigma_{ij} \tag{2.24}$$

where $E_{ij} = \partial \ln x_i / \partial \ln P_j$ holding output and the other input prices constant.

It can be also shown that in the case in which firms use a cost-minimizing behaviour, AES can be written as:

$$\sigma_{ij}^{AES} = \frac{e_{ij}}{s_j} \tag{2.25}$$

where e_{ij} is the elasticity of x_i with respect to the price w_j of input x_j and $s_j = x_j w_j / (\sum_{i=1}^{n} x_i w_i)$. In the latter expression, the nominator is the total expenditure for input x_j and the denominator is total expenditures.

AES have been intensively employed in applied literature although they have been harshly criticised by Blackorby and Russell (1989). Indeed, although an AES applied to the two inputs case returns the same value as the original Hicks's (1932) elasticity of substitution, Blackorby and Russell (1989) demonstrated that it does not share the same properties and that *"As a quantitative measure, it has no meaning; as a qualitative measure, it adds no information to that contained in the (constant output) cross-price elasticity"* (Blackorby and Russell, 1989, p.883). Hence, they revived an alternative measure conceived by Morishima (1967), MES. This informs on the percentage change in two inputs ratio given a percentage change in the price of one of the two inputs. It can be formulated as follows:

$$\sigma_{ij}^{MES} = \frac{f_j}{x_i} \frac{|D_{ij}|}{|D|} - \frac{f_j}{x_j} \frac{|D_{jj}|}{|D|} \tag{2.26}$$

This is an asymmetric measure and can be written in terms of AES as:

$$\sigma_{ij}^{MES} = \frac{f_j x_j}{f_i x_i} (\sigma_{ij}^{AES} - \sigma_{jj}^{AES}) \tag{2.27}$$

Factors that are AES substitutes are MES substitutes, factors that are AES complements might become MES substitutes. Only from 2004, applied work begun employing Morishima elasticities (see in Table 2.1, columns EK, EL, KL). Of particular interest is the work by Kim and Heo (2013) on ten OECD countries between 1980 and 2007. The authors reflect on the importance of the asymmetric substitutability between energy and capital that can be captured by MES: if energy costs increase faster than capital costs then an entrepreneur should find convenient to invest in energy-saving machineries. However, what the authors find is that, although energy prices grew more rapidly than capital prices, the degree of substitutability of energy for capital dominated the substitutability of capital for energy indicating that energy pricing does not always lead to the adoption of energy-saving technologies.

Another paper worth mentioning is the one by Frondel (2011) which showed how all elasticities can be re-written in terms of cross-price elasticities.

When we consider nested CES production functions, the implied elasticities are HES. Blackorby and Russell (1989) show that, while HES, AES and MES coincide in a 3-input unnested CES (and are constant), in a nested CES this is not always guaranteed. Indeed, HES are constant at each level of production, but MES and AES are not because they do not hold other inputs quantities fixed. It is true, however, that in nested CES functions the AES between the each nested input and the input/s outside the nest must coincide.

2.4.1 An early debate

This is the appropriate time to give a brief digression which involves net/gross and short/-long run elasticities of substitution. Berndt and Wood (1975, p. 259) described their work as the first which *"has explicitly investigated cross substitution possibilities between energy and non-energy input"*. They utilized data from the US manufacturing sector covering the period 1947-71 and, using a Translog cost function, estimated a negative elasticity of substitution between energy and capital which identified them as complements. The year after, Griffin and Gregory (1976) published an article which pooled nine countries (including the US) using cross-sectional data for the four years 1955, 1960, 1965, 1969 and came to the opposite conclusion: capital and energy are substitutes. It is important to note that, aside from the decision to opt for cross-sectional and time series data, the other difference between the two studies regards the assumption of homotheticity and weak separability: Griffin and Gregory (1976) hypothesized that the material input was weakly separable from the other inputs.

In order to reconcile their results with Berndt and Wood (1975), Griffin and Gregory (1976) proposed the distinction between long-run and short-run elasticities: models based on time-series (TS) are static, reflect short-term adjustments to price variation and presume that capital stock has no time to be technologically adjusted; therefore, TS lead to short-run elasticities of substitution. On the other hand, cross-sectional (CS) data are connected with long-run movements and not fixed technology revealing long-run elasticities. According to the authors, capital and energy can be short-run complements and long-run substitutes.

Still, Berndt and Wood (1979) did not concur with this thesis and traced the origin of the two conflicting estimation results in the difference between gross and net cross-price elasticities. In their opinion, while their original estimates represented net elasticities (which measure the ease with which capital and energy could be substituted holding output fixed), Griffin and Gregory (1976) measured a gross elasticity. In fact, excluding the material input, they were computing an elasticity which was holding constant only the output of the weakly separable subfunction (e.g. $X^m = f(K, L, E)$). They formally expressed the relationship between net and gross elasticities as:

$$E_{ij}^{Net} = E_{ij}^{Gross} s + S_{im} E_{mm} \tag{2.28}$$

where the last term is called the expansion elasticity and it comprises the cost share of the jth input in the total cost of producing X_m and E_{mm} is the own-price elasticity of demand for X_m. The authors stated that a negative expansion elasticity that exceeds the gross elasticity reconciles the substitution/complementary results.

The discussion between the two groups of authors proceeded until 1981 but did not reach a final agreement. In 1986, Anderson and Thursby (1986) provided an answer. They showed that point elasticity estimates provide less information than confidence intervals that they constructed on the ratio of normals and the normal distribution of the AES: *"examination of the confidence intervals demonstrates that point-estimates of the elasticities often provide no information regarding the structure of the technology or factor demand: the confidence intervals span both positive and negative values."* (Anderson and Thursby, 1986, p.647). They re-estimated Berndt and Wood's (1975) and Griffin and Gregory's (1976) models and found that neither capital/energy substitutability nor complementarity are supported by a 95% confidence interval about the estimated elasticity value.

Following Anderson and Thursby (1986), Hisnanick and Kyer (1995), and Medina and Vega-Cervera (2001) constructed confidence intervals for the AES and demand elasticities. The first study was performed on the US manufacturing sector in the period 1958-85 and came to the conclusion that energy and capital are substitutes. The second study, that has already been described above, confirmed for Spain, Portugal, and Italy that the sign of the relationship between the two inputs is ambiguous.

2.5 Data

2.5.1 Data aggregation

One of the main problems regarding the estimation of the elasticities of substitution regards the selection of an appropriate dataset. In this context, two matters need consideration. The first one concerns the level at which data are collected, the second regards the way factor inputs are constructed. We treat this two issues separately in the following two subsections.

2.5.1.1 Level of data aggregation

Until 1980, the literature on the estimation of substitution elasticities was based on data aggregated at the level of the manufacturing sector of the country (or countries) of interest (see Table 2.1, column Sect). This implicitly meant that the elasticities of substitutions between inputs could be considered the same at each subsector (e.g. Engineering, Textiles, and Clothes, Sheltered food or Chemicals). Eventually, Field and Grebenstein (1980) estimated different elasticities for ten two-digits manufacturing industries in the US, and found the results to vary both quantitatively and qualitatively across the different subsectors. Dargay (1983) compared the results of the entire manufacturing sector with those of twelve single industries and reported that, although in general the total manufacturing elasticities estimates maintain the same sign, the variation across specific industries is remarkable in terms of both the magnitude and the nature of the substitution responses. Therefore, Dargay (1983, p.47) underlined *"the importance of disaggregating manufacturing into its component industries.* [...] *Estimates based on total manufacturing are thus not generally representative for individual industries* [...] *as these partially reflect changes in relative production shares over the observation period"*.

An additional level of disaggregation of the US manufacturing sector is called for by Prywes (1986), who emphasized the necessity of looking at each industry in order to avoid aggregation errors. In his model, the single observation for any particular year for each two-digit SIC industry is constructed using its four-digit SIC member industries. Hazilla and Kopp (1986) and Iqbal (1986) confirmed the evidence of high intersectoral variation in the elasticities of substitution.

A different approach that is worth mentioning has been applied in the works of Nguyen and Streitwieser (1999), Arnberg and Bjorner (2007), Haller and Hyland (2014) who looked at micro-level data using as an observation the single company. In particular, they considered, respectively, 10,412 four-digit US plants, 903 Danish companies, and Irish companies. Nguyen and Streitwieser (1999) and Arnberg and Bjorner (2007) papers were based on a cross-sectional estimation and evidenced that energy and capital are substitutes, while Arnberg and Bjorner (2007), who built a panel dataset found that the estimated capital/energy elasticity of substitution was negative, indicating complementarity.[20]

Undoubtedly, the framework in which the industrial disaggregation is of greatest interest is in general equilibrium modelling. In that context, disaggregation allows a nearer approximation to the real economy. This explains why all the studies where the estimation was justified by the necessity of empirical foundations for policy modelling, the manufacturing sector was broken up into several sub-sectors. In general, it emerged that disaggregation allows for a more precise simulation of each industrial sector since the outcomes are very diverse in what regards the elasticities magnitude and, as it was explained earlier, the nested structures of the CES production function.

2.5.1.2 Input aggregation

The research on inputs substitutability in the last forty years has focused mainly on aggregate capital and energy, with the aim of shedding light on the nature of their relationship, given the conflicting evidence found in the earliest applied works. But capital and energy, like labour and material inputs, are themselves aggregates, in the sense that all of them are made of different components. For instance, the energy aggregate comprises different components, e.g. natural gas, electricity and oil, and the capital input can be broken apart into machines and structures. With this in mind, part of the literature has investigated if the knot could be unravelled by looking at how the substitutability between two factors varies within the different input components.

Fuss (1977), Pindyck (1979), Turnovsky et al. (1982), and Iqbal's (1986) work were the first to be motivated by the idea of revealing the importance of disaggregating the energy input. They estimated sub-models where they broke apart the energy input into four or six components with the aim of estimating interfuels substitutability. In light of those cited studies, Hesse and Tarkka (1986), Ilmakunnas and Torma (1989), Arnberg and Bjorner (2007), and Kim and Heo (2013) split energy into electricity and fuels. Only Ilmakunnas and Torma (1989) found that, under certain conditions, fuel and energy have an opposite relationship with capital; the other authors showed that the two components have the same relation to the capital aggregate. Moghimzadeh and Kymn (1986) and Hisnanick and Kyer

[20]Arnberg and Bjorner (2007) noted that a linear logit model appears to be more appropriate than the Translog model when micro-data are employed.

(1995) performed separability tests on a five inputs translog cost function (considering capital, labour, electric energy, non-electric energy, and material) for the US during the periods 1954-77 and 1958-85 respectively, and both maintained that the electric and non electric partition is statistically justified. However, two different conclusions were reached: Moghimzadeh and Kymn (1986) found capital and electricity to be complements and capital and non-energy inputs to be substitutes, while Hisnanick and Kyer (1995) stated that both the components and capital were substitutes in production.

Three studies for the US disaggregated the capital input. Field and Grebenstein (1980) divided capital into working and physical capital and found working capital and energy to be substituted, and physical capital and energy to be complements. Hazilla and Kopp's (1986) capital components are structures and equipment and the estimates reported in both cases a substitution relationship with energy. Garofalo and Malhotra (1988) is the only study which performed a weak separability test on the two components (buildings and equipment) and found statistical support for capital disaggregation. They found a negative elasticity of substitution between building and energy, and a positive one between machinery and energy.

2.5.2 Measurement issues

Depending on the functional form specified, different data are needed. For a Translog cost function model estimation, the price of inputs and the relative cost shares need to be collected; for a CES production function, only input quantities are required. Before 2000, gathering data for a single country was demanding, but creating a dataset for a pool of different nations was almost impossible. Authors had to deal with multiple national sources and accounts and this noticeably increased the probability of measurement errors. The problems connected with data measurement regarded mostly the way data on inputs were aggregated, especially capital. It was often the case that authors had to build their own measures by, for instance, employing Divisia quantity and price indexes.

For labour quantity and price, Berndt and Wood (1975) and Hisnanick and Kyer (1995) built a Divisia index in man-hours and when computing P_L, they just divided the total labour compensation by the quantity index. Other authors used, in general, the number of employees in man-hours as quantity and the average wage or hourly wage as price.

Material input represented a real challenge as data were not readily available. For this reason, most studies do not include intermediate goods. Among the nine papers whose production/cost function was based on KLEM inputs and that provided information on data-gathering (see Table 2.1, column M), few constructed Divisia indexes and the rest defined non-energy inputs as quantity and the relative average cost as price.

The quantity of energy was usually measured through an index of energy consumption reported in several different measurement units and P_E was either an average of the total expenditure or a price Divisia index. Fuss (1977), in his study on five Canadian regions, derived the energy price endogenously using the parameter estimates of a submodel on energy components and demonstrated that it behaves better then a Divisia index.

Capital deserves a special discussion as its measurement has always been troublesome due to the complexity of reconciling the theory with the empirical data. According to production theory, the quantity of capital input is represented by the flow of services provided by the capital goods. However, there are no readily available measures of the flow of capital services which is, in fact, a very abstract concept: it includes all the explicit and implicit transactions connected with capital goods in each production period. Hence, if the firm owns a particular capital asset such as a machine, the rental price or user cost for this asset in each period is implicit and does not appear in the accountancy books. On the same line, a machine is usually deployed for more than one period but the explicit transaction cost appears only in the accountancy year in which it has been purchased. Neoclassical theory[21] has linked the quantity of capital services to the quantity of capital goods[22] (stock of capital) defining the quantity of services as a measure of the contribution of the capital stock to the production process in a given year. The capital stock is an aggregate that can include several types of goods such as equipment and structures, intangibles (e.g. software), land, financial assets and human capital. However, National Accounts traditionally exclude the last two from the capital stock. In order to estimate capital services, the latest approach used by the OECD is the following: calculate the net stock series from investment series using a perpetual inventory model which accounts for age-efficiency profile and depreciation patterns, then estimate the rental price of each asset (that is the cost of the asset for one period) and that gives back the price of the capital services; finally, use these two steps to generate weights for each input component and combine them. Despite these indications, recent literature have generally used net, gross or fixed capital stock instead of a computed measure of capital services.[23]

Field and Grebenstein (1980) distinguished between two approaches used in estimating the cost of capital: the value added and the service price methods. The first one was used

[21] Jorgenson and Griliches (1967), Hall and Jorgenson (1967) and Hulten (1990).

[22] Jorgenson and Griliches (1967) proposed the idea of capacity utilization but it has been demonstrated that this entailed ulterior measurement problems.

[23] Gross capital stock after depreciation is net capital stock.

by Griffin and Gregory (1976), Fuss (1977), and Pindyck (1979) who derived the price as the difference between value added and payroll. The second, used by Berndt and Wood (1975), multiplies the rental price of capital services for the physical capital. Field and Grebenstein (1980) showed that the service price method yields to K-E complementarity and the value added approach to K-E substitutability, providing evidence that the way capital is measured influences the final substitution estimates. Finally, Hazilla and Kopp (1986) demonstrated that different service price specifications lead to statistically different elasticities estimations using 34 alternative definition of capital service price.

From 2000, a few European and international database were introduced by the European Commission and the OECD such as the EU-KLEM database and the WIOD database. As a consequence, almost all the work published after 2000 are based on a panel framework.

2.6 Econometric techniques

2.6.1 Translog cost function

When the selected functional form is Translog and the duality theorem is exploited in order to take advantage of the facilitating characteristics of the cost function, the estimation procedure reduces to a system of linear equations. These may be subject to the restrictions imposed by the assumptions of homotheticity and, in certain cases, linear homogeneity and separability. Indeed, in order to obtain an estimate of the AES between two inputs, one needs to estimate the parameters of the demand functions.

The most common estimation technique involves appending a stochastic additive error to each cost share equation as follows:

$$S_i = b_i + b_{ii}\ln(P_K) + \sum_{j=1}^{n} b_{ij}\ln(P_j) + b_{Kq} + b_{it}t\ln(q) + \epsilon_i \qquad \text{with } i, j = 1, ..., n \quad (2.29)$$

The disturbance terms represent both random errors in the cost-minimizing behaviour and random influence of omitted variables. Since the sum of the four share equations equals one, the disturbance covariance matrix is singular and non-diagonal. The approach used in the literature to overcome this problem is to drop arbitrarily one equation from the system so that the resulting vector of disturbances is composed of identically and independent normally distributed error terms with mean zero and a non-singular covariance matrix. This allows for the correlation between contemporaneous errors of different equations to be nonzero. In order to obtain consistent and asymptotically efficient estimates, the chosen estimation technique must be invariant to the equation deleted. Two possible and asymptotically equivalent procedures have been employed in the literature: an iterative[24] version of the Seemingly Unrelated Equations (ISUE) regression by Zellner (1962) and the Full Information Maximum Likelihood (FIML) estimation procedure. In all Translog studies, ISUE or FIML estimators are applied with the price homogeneity and symmetry parameter restrictions imposed (see Table 2.1, column Ec).

Few papers included the cost function in the system of estimated equation.[25] This allows the authors to test for constant returns and Hicks neutrality. However, it implies the estimation of a cost function composed of a large number of coefficients which may lead to multicollinearity. As a consequence, standard errors may be large and coefficients difficult to interpret.

Eight out of the thirty-one studies which used the Translog function worked on time series-cross sectional data of pooled countries or sectors (see Table 2.1, column Str). Different models were adopted: the basic one estimates a system of equations where the parameters are assumed to be the same for each country (sector); an intermediate model where the b_i parameters are country-specific (sector-specific); the more complex one where all the parameters are allowed to vary across countries (sectors), and thus it implies estimating a system of equation for each country considered. Griffin and Gregory (1976) compared the goodness of fit of the three models using the R^2 statistic and found the second to explain better the data. However, they argue that, as long as the parameter estimates of the slopes do not vary noticeably, the first model should be preferred because in this way the inter-country mean variation is not eliminated. Pindyck and Rotemberg (1983) compared the same models through a χ^2 test and found that different intercepts across countries should be allowed. Finally, Fuss (1977), Ozatalay et al. (1979), Hesse and Tarkka (1986), Iqbal (1986), Garofalo and Malhotra (1988), and Roy et al. (2006) introduced country dummy variables in the cost shares and tested for their significance.

Three special cases are represented by Arnberg and Bjorner (2007), Haller and Hyland (2014), and Khiabani and Hasani (2010) who, in their micro-estimation, used a fixed effect estimator to account for the panel nature of the data and Christopoulos (2000) who specified a dynamic model based on first differences after testing for unit roots.

[24]Until the estimated coefficients and residuals covariance matrix converge.

[25]Norsworthy and Malmquist (1983), Nguyen and Streitwieser (1999), Burki and Khan (2004), Khiabani and Hasani (2010), Haller and Hyland (2014), Zha and Ding (2014), Zha and Zhou (2014).

2.6.2 CES production function

Five different estimation techniques have been employed with a CES model and three of them involve the resolution of a cost minimization problem.

According to Prywes (1986) and Chang (1994), who used a three-level nested structure, a cost minimizing procedure needs to be employed at each of the three nests (K, E; KE, L; KEL, M) starting with the inner one. For instance, the inner nest minimization can be specified as follows:

$$\min_{KE} P_K K + P_E E \tag{2.30}$$

$$\text{subject to:} \qquad q_{KE} = \left(\beta(K)^{\frac{-1+\sigma_{K,E}}{\sigma_{K,E}}} + (1-\beta)(E)^{\frac{-1+\sigma_{K,E}}{\sigma_{K,E}}}\right)^{\frac{\sigma_{K,E}}{-1+\sigma_{K,E}}} \tag{2.31}$$

where r and f are the rental cost of capital and the price of energy, and q_{KE} is the intermediate output. Solving the minimization problem, a FOC is derived:

$$\frac{K}{E} = \left(\frac{\beta}{(1-\beta)}\right)^{\sigma_{K,E}} \left(\frac{P_E}{P_K}\right)^{\sigma_{K,E}} \tag{2.32}$$

As they assumed exogenous prices, in the next step they equated the unit cost of q_{KE}, that is P_{KE}, to the Lagrangian multiplier or shadow price. Finally, adding a disturbance term, they estimated the logarithm of equation (2.32). This procedure was repeated in the two upper levels of production using each time, as one of the inputs, the intermediate input computed from the estimated coefficients of the previous level. The elasticities of substitution are, therefore, represented by the coefficients attached to the logarithm of the ratio between prices and the share parameters can be derived from the constant term. This method can be used also with increasing or decreasing returns to scale but with the limit that it would not be possible to disentangle the share parameter from the scale parameter.

Differently, Kemfert (1998) and Koesler and Schymura (2015) employed a direct method by estimating three non-linear equations for each nested structure selected.

Recently, van der Werf (2008) followed an indirect method closely related to the first one presented. He minimized a cost function at each nest and found input demand equations. However, since he considered factor-specific technology parameters, his conditional input demand equations were under-identified. Hence, he took first differences and after few algebraic steps, he ended up with four equations for each nested structure to which he added an error term. He, then, employed a fixed effect estimator where the within variation was due to dummy variables constructed for each country-industry combination.

Baccianti (2013), who also had to face the problem of under-identification, proposed a new approach based on a panel normalization procedure to identify the input demand equations for twenty-seven countries. He estimated the normalized equations using a generalized method of moment estimator with a variance-covariance matrix robust to heteroskedasticity and autocorrelation.

2.7 Economic context

The last explanation of the contradictory results on the substitution elasticity between capital and energy is provided by the different context in which they have been estimated. An economic context is defined by a geographic area and a time period. Concerning the former, the early literature has been mainly focused on the US and Europe (see Table 2.1, column Country). Koetse et al. (2008), in their meta-analysis, tested for the difference in the estimates between these two regions and concluded that this is substantial. Recently, a number of papers has focused on China given its great expansion largely sustained by energy consumption.

Referring to the time period, empirical work can be divided into two periods: pre-oil crisis or post-oil crisis. Three different analyses have been proposed in order to check if the estimation period has an effect on final elasticities. The first is by Hesse and Tarkka (1986) who estimated the same model for the periods before and after the crisis and found that the price sensibility of demand for inputs has been influenced by the change in the price regime. A second work is due to Ilmakunnas and Torma (1989) who estimated a model in which they verified the presence of a change in the structural parameters: in the period 1960-73 they found energy-capital complementarity while in the period 1974-1981 they observed energy-capital substitutability. Lastly, Koetse et al. (2008) found small differences in the estimates for the two periods.

2.8 Conclusions

In this literature review, we discussed forty works on the econometric estimation of the elasticity of substitution between energy and other inputs, spanning four decades. What emerges is that, regardless of the abundance of papers and the use of novel techniques and appropriate datasets, findings are discordant: for example, there is no consensus on the nature of the relationship between capital and energy. This literature review is structured around five main aspects that, in our opinion, justify this variety in the results: the assumptions on the production function, the type of elasticity, the level of data collection, the estimation techniques and the economic context.

While it is understandable that different data and techniques would lead to different results, one would expect that the same dataset and the same aim would demand the same approach. However, we find that this has not always been the case, and that the choices made on the previously mentioned aspects are rarely justified. This calls for a uniform and solid procedure that should guide researchers in the estimation. We recommend that this includes the following steps.

The first step, that has been overlooked so far but is crucial given the new available databases, is to run diagnostic and formal tests on the data. In particular, stationarity of the time-series of prices or quantities should be checked as well as cointegration. Furthermore, with panels of industries followed over several years, not only serial correlation but also simultaneous correlation of the error term should be tested as these are generally characterized by a number of sectors that is bigger than the number of yearly observations available.

The second step is represented by formal tests on the desired assumptions: in the chapter we pointed out that restrictions on inputs and technology are often assumed albeit, in most of the cases in which they were tested, they were rejected.

The third step concerns the appropriate estimation technique. In the existing literature, the main econometric procedures adopted are the Full Information Maximum Likelihood and the Iterative Seemingly Unrelated Regressions. However, new estimators are now available for both panel data and cross-sectional data that do not require such strong distributional assumptions. Indeed, ISURE estimates are not consistent in the presence of serial correlation or heteroskedasticity which are often an issue with input-price series.

Finally, as the fourth step we recommend to choose *a priori* which type of elasticity of substitution to compute as they have different interpretations. When using nested CES functions, only Hicks elasticities are obtainable. Allen elasticities have been extensively employed in the past, and thus calculating them permits comparison with previous studies. Morishima elasticities deliver more information: being asymmetric they allow to look at two degrees of substitutability for every pair of inputs.

Future research should acknowledge these steps and, using the latest available dataset (i.e. EU-KLEM and WIOD), produce a new set of results based on a common methodology.

This will simplify the comparison between country estimates and shed light on the real nature of the relationship between energy and capital.

Furthermore, we warn researcher attempting the estimation of elasticities of substitution to inform CGE models to be careful in employing nested CES functional forms. Although so far this choice has been driven by practical needs (i.e. CGE models are traditionally based on these production function), it should instead be empirically supported by the dataset under analysis.

Table 2.1: A summary of the literature in chronological order

Author	Fcn	HT	CRTS	S	HN	M	Disag	Time	Sect	Str	Ec	Country	EK	EL	KL
Hudson and Jorgenson (1974)	TL	x	x	x		x		1947-71	9 SEC	TS	SUE	US	-1.37	2.16	1.09
Berndt and Wood (1975)	TL	x	x	T	x	x		1947-71	MS	TS	ISUE	US	-3.22	0.64	1.01
Griffin and Gregory (1976)	TL	x	x	T	x			1955-69	MS	P-CS-TS	ISUE	US 8 EU	1.07 1.03	0.87 0.83	0.06 0.43
Fuss (1977)	TL	T		x	x	x	E(6)	1961-71	MS	P-CS-TS	ISUE NIV	5 CA	-0.05^{CP}	0.55^{CP}	0.20^{CP}
Pindyck (1979)	TL	T		x	x		E(4)	1957-73	MS	P-CS-TS	ISUE	US 7 EU JP CA	1.77 0.58 0.74 1.48	0.05 1.14 1.15 0.42	1.41 0.7 0.70 1.43
Danny et al. (1978)	GL	T			T	x		1947-70	MS	TS	ISUE	CA	-11.91	4.89	5.46
Berndt and Wood (1979)	TL	x	x		x	x		1971	MS	TS	ML	US	-0.33	-	-
Magnus (1979)	GCD	x	x					1950-76	MS	TS	FIML	NL	-2.32	1.25	0.89
Ozatalay et al. (1979)	TL	x	x		x	x		1963-74	MS	P-CS-TS	ML	US WDE JP	1.22 1.15 1.18	1.03 1.04 1.05	1.08 1.06 1.14
Field and Grebenstein (1980)	TL	x	x	x	x		K(w,p)	1971	10 SEC	CS	ISUE	US	w=2.09 p=-3.18	0.07	w=0.34 p=0.25
Turnovsky et al. (1982)	TL	x	x	x		x	E(4)	1946-75	MS	TS	FIML	AT	2.26	-2.66	2.00
Dargay (1983)	TL	T			x	x		1952-76	12 SEC	TS	FIML	SE	-1.43	0.12	0.66
Norsworthy and Malmquist (1983)	TL	x	x	T	T	x		1969 1977	MS	TS	ISUE	JP US	-13.06 -14.57	- -	- -
Hazilla and Kopp (1986)	TL	x	x	T		x	K(e,s)	1958-74	21 SEC	TS	ISUE	US	Ke=1.70 Ks=7.60	- -	Ke=0.30 Ks=-0.70
Hesse and Tarkka (1986)	TL	x	x	x	T		E(e,f)	1960-73	MS	P-CS-TS	FIML	8 EU	e=-0.38 f=1.29	e=0.66 f=0.49	1.20
Iqbal (1986)	TL	T	T	x	x			1960-70	16 SEC	P-CS-TS	ISUE	PK	1.64	-0.50	0.88
Moghimzadeh and Kymn (1986)	TL	x	x	T	x	x	E(e,ne)	1954-77	MS	TS	ISUE	US	e=-0.06 ne=0.10	e=0.06 ne=-0.16	0.27

Table 2.1 – *Continued from previous page*

Author	Fcn	HT	CRTS	S	HN	M	Disag	Time	Sect	Str	Ec	Country	EK	EL	KL
Prywes (1986)	CES	x	x	x	x	x		1971-76	450 IND	P-CS-TS	OLS	US	-1.35	0.88	0.88
Garofalo and Malhotra (1988)	TL	x	x	T	T		K(b,m)	1963-66 1974-78	MS	P-CS-TS	ISUE	US	-0.70	2.41	0.72
Ilmakunnas and Torma (1989)	GL	x	x		x	x	E(e,f)	1960-73 1974-81	MS	TS	NLML	FI	e =-/0.45 f =-/-0.73	e=1.16/0.55 f =-0.85/-	0.24/0.36
Chang (1994)	CES	x	x	x	x	x		1956-71	MS	TS	OLS	TW	2.17	0.35	0.35
Hisnanick and Kyer (1995)	TL	T		T	x	x	E(e,ne)	1958-85	MS	TS	ISUE	US	e=0.31 ne=0.95	e=-5.03 ne=-16.80	0.58
Kemfert (1998)	CES	x	x	x				1960-93	MS	TS	NLE	DE	0.65^{H}	0.84^{H}	0.84^{H}
Nguyen and Streitwieser (1999)	TL	x	x		x	x		1991	10,412 IND	CS	ISUE	US	0.58	4.01	0.002
Christopoulos (2000)	TL	T		x	x			1970-90	MS	TS	NLISUE	GR	0.25	0.05	0.33
Medina and Vega-Cervera (2001)	TL	T		T	x			1980-96	MS	TS	FIML	IT PT ES	0.10 0.17 -0.01	0.12 0.15 0.95	0.43 -0.09 0.30
Burki and Khan (2004)	TL	T			T	x		1969-90	MS	TS	ISUE	PK	0.07	1.51	2.92
Roy et al. (2006)	TL	x	x	T	T	x		1980-93	MS	P-CS-TS	ISUE	IN KR BR	4.48 7.68 9.80	-1.76 -2.03 -0.36	5.11 5.19 9.47
Arnberg and Bjorner (2007)	TL	T		x	T		E(el,oe)	1993-97	80 IND	Pa	FE	DK	el=-0.20^{CP} eo=-0.24^{CP}	el=0.37^{CP} eo=0.62^{CP}	0.56^{CP}
Okagawa and Ban (2008)	CES	x	x	x				1995-04	19 SEC	Pa	FE	14 OECD	0.21^{H}	0.33^{H}	0.33^{H}
van der Werf (2008)	CES	x	x	x				1978-96	7 SEC	Pa	OLS	12 OECD US	0.35^{H} 0.54^{H}	0.35^{H} 0.54^{H}	0.41^{H} 0.32^{H}
Khiabani and Hasani (2010)	TL	T	T	x	T			1984-05	9 SEC	Pa	ISUE	IR	-0.61	2.15	2.81
Dissou et al. (2015)	CES	x	x	x	T			1962-97	10 SEC	TS	SUE	CA	0.29^{H}	0.82^{H}	0.29^{H}
Koesler and Schymura (2015)	CES	x	x	x		x		1995-06	35 SEC	Pa	NL	40 OECD	0.78^{H}	0.78^{H}	0.49^{H}
Ha et al. (2012)	CES	x	x	x				1970-05 1961-05	27 SEC 35 SEC	Pa	OLS	GB US	0.31^{H} 0.14^{H}	0.18^{H} 0.16^{H}	0.31^{H} 0.14^{H}

Table 2.1 – *Continued from previous page*

Author	**Fcn**	**HT**	**CRTS**	**S**	**HN**	**M**	**Disag**	**Time**	**Sect**	**Str**	**Ec**	**Country**	**EK**	**EL**	**KL**
Baccianti (2013)	CES	x	x	x				1995-08	33 SEC	Pa	GMM	27 OECD	0.57^{H}	0.38^{H}	0.38^{H}
Kim and Heo (2013)	TL	T			T		E(e,f) K(2)	1980-07	MS	TS	ISUE	10 OECD	e=0.08^{M} f=0.22^{M}	e=-0.17^{M} f=-0.78^{M}	0.29^{M}
Haller and Hyland (2014)	TL	T	T		x	x		1991-09	IND	Pa	ISUE	IE	1.54^{M}	0.61^{M}	0.61^{M}
Zha and Ding (2014)	TL	T		x	x			1994-08	MS	TS	FIML	CN	2.10	-0.05	-
Zha and Zhou (2014)	TL	T		x	x			1995-08	1 SEC	TS	FIML	CN	0.67	0.25	1.19

Fcn: Functional Form. CES: nested Constant Elasticity of Substitution, GCD: Generalized Cobb Douglas, GL: Generalized Leontief, TL: Translog;
HG: Linearly homogeneous, S: Weakly separable, HT: Homothetic, HN: Hicks-neutral, M: input M is considered. T stands for tested;
Disag: Input disaggregation;
Sect: Sectors. SEC: Sector, IND: Industry, MS: Manufacturing Sector;
Str: Structure of the Data. CS: Cross sectional, P: Pooled, Pa: Panel, TS: Time-series;
Ec: Econometric Technique. FE: Fixed Effect, FIML: Full Iterated Maximum Likelihood, GMM: Generalized Method of Moments, ISUE: Iterated Seemingly Unrelated Equations, ML: Maximum Likelihood, NIV: Non Linear Instrumental Variables Estimator, NL: Non Linear Estimator, NLISUE: Non Linear Iterated Seemingly Unrelated Equations, NLML: Non Linear Maximum Likelihood, OLS: Ordinary Least Squares, RIDGE: Ridge Estimator, SUE: Seemingly Unrelated Equations;
Country: ISO 3166-1 alpha-2 Country Code;
EK, EL and KL: Allen elasticities if not specified, CP superscript if Cross Price Elasticities, H superscript if Hicks elasticities, and M superscript if Morishima Elasticities.

Chapter 3

On Translog Separability and the Linear Approximation of Nested CES

3.1 Introduction

Empirical studies often assume the production function to be separable in their inputs for three main reasons. Firstly, if a production function is separable in its inputs, the decision making process concerning the optimal input quantities can be tackled in subsequent steps. For instance, if the production technology is based on capital, energy and labour, separability of capital and energy implies that the producer can first optimize the intensity of the so-called "utilised capital" intermediate factor and then find the relative efficient quantity of labour. Secondly, separability justifies the use of aggregated inputs that is typical of applied works (e.g. capital aggregation) and also the value-added measures of output. Lastly, when data availability implies discarding one of the inputs, separability guarantees that production efficiency is not affected.

Functional separability was at first defined and explored for consumption theory by Strotz (1959) in order to partition the utility function in subsets of commodities and form the so-called "utility tree". The seminal paper by Sato (1967) translated this definition in production terms for the purpose of defining a new category of production functions, namely the two-level Nested Constant Elasticity of Substitution (CES). Thereafter, Berndt and Christensen (1973a,b) explored input separability in the particular case of a Translog production function. They provided the expression for the constraints that a researcher needs to impose on the Translog coefficients to attain separability and an example of a three-inputs Translog. In the subsequent years, a number of papers exploited those constraints to verify the assumption of separability for their datasets. (Berndt and Wood, 1975, Hazilla and Kopp, 1986, Garofalo and Malhotra, 1988, Hisnanick and Kyer, 1995, Medina and Vega-Cervera, 2001, Roy et al., 2006)

The main contribution of this chapter is to generalize Berndt and Christensen's (1973b) analysis to the n-input Translog clarifying the number of constraints needed and how to express them. The existing literature has so far only focused on particular separability structures involving at most pairs of inputs, we intend to provide guidelines on how to deal with more complex forms of separability. Indeed, when considering more than three inputs, a naïve application of the separability definition leads to the imposition of a large number of constraints that drastically increases with the number of inputs. Nevertheless, some of these are not linearly independent from the others and can therefore be ignored. For this purpose, we show an approach that can be employed to identify the necessary and sufficient constraints. This approach is based on the comparison between nested CES and Translog production functions by means of a linear approximation of the former. The second-order multivariate Taylor expansion of nested CES functions has never been attempted so far, thus it represents a further contribution of this chapter.

The structure of the chapter is the following. First, the definition of separability is investigated for a general production function and for the Translog, highlighting in particular which ones are its main limits and drawbacks. Second, we illustrate a method that can be used to overcome the aforementioned limits: a general rule producing the number of required constraints and an approach for the identification of the constraints of interest. Finally, conclusions are drawn.

3.2 Berndt and Christensen's (1973a) definition of functional separability

Berndt and Christensen (1973a) provided the following definition of weak and strong separability for production inputs:[1]

> We consider a twice-differentiable, strictly quasi-concave homothetic production function with a finite number of inputs, each having a strictly positive marginal product. We denote this production function $Q = F(x) = F(x_1, \ldots, x_n)$. The set of n inputs is denoted $N = [1, \ldots, n]$, and is partitioned into r mutually exclusive and exhaustive subsets $[N_1, \ldots, N_r]$, a partition which we shall call R.
>
> The production function $F(x)$ is said to be weakly separable with respect to the partition R if the marginal rate of substitution (MRS) between any two inputs x_i and x_j from any subset N_s, $s = 1, ..., r$, is independent of the quantities of

[1] Note that, in the following quote, we change the notation to reflect the one used in the remaining of this chapter.

inputs outside of N_s, i.e.

$$\frac{\partial}{\partial x_k}\left(\frac{F_i}{F_j}\right) = 0 \quad \text{for all} \quad i, j \in N_s, \quad k \notin N_s. \tag{3.1}$$

where F_i represents the partial derivative of $F(x)$ with respect to input x_i. The production function $F(x)$ is said to be strongly separable with respect to the partition R if the MRS between any two inputs from subsets N_s and N_t, does not depend on the quantities of inputs outside of N_s and N_t, i.e.

$$\frac{\partial}{\partial x_k}\left(\frac{F_i}{F_j}\right) = 0 \quad \text{for all} \quad i \in N_s, \quad j \in N_t \quad k \notin N_s \cup N_t. \tag{3.2}$$

While condition (3.2) always implies condition (3.1), the opposite is only true when there are only two subsets.

The authors showed that the separability conditions (3.1) and (3.2) can be summarized by:

$$F_j F_{ik} - F_i F_{ik} = 0 \quad \text{for all} \quad i, j \in N_s, \quad k \notin N_s \tag{3.3a}$$

$$F_j F_{ik} - F_i F_{ik} = 0 \quad \text{for all} \quad i \in N_s, \quad j \in N_t \quad k \notin N_s \cup N_t. \tag{3.3b}$$

Equation (3.3a) and (3.3b) refers to weak and strong separability respectively.

Weak separability is a necessary and sufficient condition for $F(x)$ to be written as $F(X_1, ..., X_r)$ where X_s is a function of the element of N_s only. Strong separability (or additive separability) is a necessary and sufficient condition for $F(x)$ to be written as $F(X_1 + \cdots + X_r)$.

Furthermore, Berndt and Christensen (1973a) showed the existence of a link between separability and the Allen elasticities of substitution (AES) between inputs. Separability is a necessary and sufficient condition for:

$$\sigma_{ik}^{AES} = \sigma_{jk}^{AES} \quad \text{for all} \quad i, j \in N_s, \quad k \notin N_s \tag{3.4a}$$

$$\sigma_{ik}^{AES} = \sigma_{jk}^{AES} \quad \text{for all} \quad i \in N_s, \quad j \in N_t \quad k \notin N_s \cup N_t \tag{3.4b}$$

where σ^{AES} is the Allen elasticity of substitution. Equation (3.4a) refers to weak separability and (3.4b) to strong separability.

In a subsequent paper, Berndt and Christensen (1973b) showed how to apply the separability constraints to a Translog function. In particular, they considered the following Translog, characterized by inputs symmetry, homogeneity and Hicks neutrality:

$$\ln Q = \ln a_0 + \sum_{i=1}^{n} a_i \ln x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \ln x_i \ln x_j. \tag{3.5}$$

To test if inputs x_i and x_j are separable from a given input x_k, they use (3.3) and obtain:

$$a_i a_{jk} - a_j a_{ik} - \sum_{m=1}^{n} (a_{im} a_{jk} - a_{jm} a_{ik}) \ln(x_m) = 0. \tag{3.6}$$

Equation (3.6) holds if two sets of constraints are jointly satisfied.[2] Let us call them

$$C1 : \quad a_i a_{jk} - a_j a_{ik} = 0 \tag{3.7a}$$

$$C2 : \quad a_{im} a_{jk} - a_{jm} a_{ik} = 0 \qquad \text{with} \qquad m = 1, \dots, n. \tag{3.7b}$$

Obviously, this procedure must be repeated for each separability assumption one wants to test (i.e. for any triplet of inputs).[3]

Berndt and Christensen (1973b) coined two terms regarding separability: linear and non-linear separability. A Translog function is linearly separable when a_{ik} and a_{jk} are jointly null; a Translog function is non-linearly separable when (3.7a) and (3.7b) are jointly true. Linear separability implies non-linear separability.

A few years later, Blackorby et al. (1977) argued that the Translog function is "separability-unflexible": once any form of separability is imposed, the production function loses its ability of approximating any arbitrary separable production function at any given point.

Following their intuition, Denny and Fuss (1977) pointed out that separability tests depend on the interpretation the researcher gives to the Translog function. Indeed, the Translog can be seen as an exact production function or as a second-order approximation to an arbitrary production function. The authors claimed that Berndt and Christensen's (1973b) separability definition applies only to the cases in which the Translog is interpreted as the true underlying production function; on the contrary, less restrictive assumptions are needed if the Translog is an approximation: for example, only the $C1$ constraints need to be satisfied for weak separability.

[2] In order to have separability restrictions that are independent of all x_m the terms in brackets must be set to 0.

[3] In the three-input case the number of $C1$ constraints is 1. With n-inputs, this is always greater than 1.

Denny and Fuss (1977) also provided a test that can be used to see whether the Translog is the approximation to an unnested CES of the form $Q = \lambda(\delta_1 x_1^{-\rho} + \delta_2 x_2^{-\rho} + (1-\delta_1-\delta_2)x_3^{-\rho})^{-\frac{1}{\rho}}$. The test is based on a set of four constraints to be jointly verified:

$$\sum_{i=1}^{3} a_i = 1 \tag{3.8a}$$

$$\sum_{j=1}^{3} a_{ij} = 0, \quad \text{with} \quad i = 1,2,3 \tag{3.8b}$$

$$a_1 a_{23} = a_2 a_{13} \tag{3.8c}$$

$$a_1 a_{23} = a_3 a_{12} \tag{3.8d}$$

where (3.8a) and (3.8a) represent homogeneity and (3.8c) and (3.8d) represent strong separability.

The subsequent literature has mostly employed Berndt and Christensen's (1973a) definition, with the exception of Hazilla and Kopp (1986) who used Denny and Fuss (1977) test of approximate weak separability.

3.2.1 Limits of Berndt and Christensen's (1973b) method

Berndt and Christensen (1973b) proposed an example of a three-input Translog of the form $F(x) = F(G(x_i, x_j), x_k) = F(X_1, x_k)$. They were looking at the case in which production works on two levels: a first one where the subset is produced and a second where the resulting intermediate input is combined with the third input. The subsequent applied literature only considered at most two levels of production and only subsets composed of no more than two inputs. However, empirical and theoretical studies have stressed that the production technology may often be based on more than two levels of production (e.g. $F(x) = F(G(H(x_1, x_2), x_3), x_4) = F(X_2(X_1), x_4)$): for instance, it is common to have capital and energy forming an inner level that is then combined with labour, this, in turn, represents another composite input that is finally aggregated with the materials input. Moreover, it is sensible to imagine that in some industries capital, energy and labour are used at the same level of production and later on combined with intermediate materials (e.g. $F(x) = F(G(x_1, x_2, x_3), x_4) = F(X_1, x_4)$).

Unfortunately, the naïve application of Berndt and Christensen's (1973b) methodology to these more realistic cases produces a daunting number of constraints. As the separability definitions always refer to pairs of input, independently of the number of inputs inside the subsets, we need to impose that each pair of them is separable from each input outside the subset. Clearly, the number of constraints to impose increases exponentially with the number of inputs inside and outside the subset(s).

For example, let us consider a four-input Translog $F(x) = F(x_1, x_2, x_3, x_4)$ and assume that we want to test whether it is possible to write it as $F(x) = F(G(x_1, x_2, x_3), x_4) = F(X_1, x_4)$. According to Berndt and Christensen's (1973b) method, we need to impose the following twenty-four constraints:

$$
\begin{aligned}
C1: \quad & a_1 a_{23} - a_2 a_{13} = 0 \\
& a_1 a_{23} - a_3 a_{12} = 0 && && \text{(3.9a)} \\
& a_2 a_{13} - a_3 a_{12} = 0 \\
& a_1 a_{24} - a_2 a_{14} = 0 \\
& a_1 a_{34} - a_3 a_{14} = 0 && && \text{(3.9b)} \\
& a_2 a_{34} - a_3 a_{24} = 0 \\
C2: \quad & a_{1m} a_{23} - a_{2m} a_{13} = 0 && \text{with} \quad m = 1, 2, 4 \\
& a_{1m} a_{23} - a_{3m} a_{12} = 0 && \text{with} \quad m = 2, 3, 4 && \text{(3.9c)} \\
& a_{2m} a_{13} - a_{3m} a_{12} = 0 && \text{with} \quad m = 2, 3, 4 \\
& a_{1m} a_{24} - a_{2m} a_{14} = 0 && \text{with} \quad m = 1, 2, 3 \\
& a_{1m} a_{34} - a_{3m} a_{14} = 0 && \text{with} \quad m = 1, 2, 3 && \text{(3.9d)} \\
& a_{2m} a_{34} - a_{3m} a_{24} = 0 && \text{with} \quad m = 1, 2, 3
\end{aligned}
$$

The first three constraints in $C1$ (3.9a) and the first three in $C2$ (3.9c) refer to the strong separability assumption concerning the three-input X_1 subset,[4] the latter three constraints in $C1$ (3.9b) and the latter three in $C2$ (3.9d) refer to the weak separability assumption concerning the X_1 subset and the remaining input x_4.[5]

Obviously, this large number of constraints greatly reduces the degrees of freedom[6] and the statistical power of empirical tests performed on the Translog estimated coefficients. However, it can be shown that some of these constraints are not linearly independent. In

[4]Whilst not obvious, whenever the production technology is characterized by a subset including more than two inputs, each input in the subset must be seen as an aggregate that forms a partition by itself. As a consequence, strong separability must be tested for each pair of inputs included in the subset so that there is pairwise equality of all the corresponding AES.

[5]Note that C2 should have been composed by eight rather than six constraints. However when $m = k$ it is immediate to see that the constraint $a_{ik} a_{jk} - a_{jk} a_{ik} = 0$ is trivially satisfied.

[6]This is especially relevant in real data applications as the number of observations on input quantities for single industries or for countries manufacturing sector is still limited.

the specific example above, only two out of three constraints in each group are linearly independent: thus, only sixteen out of the twenty-four constraints are necessary and sufficient. Identifying how many and which sufficient constraints are relevant when the number of inputs increases becomes more and more elaborate. A possible approach to overcome this issue is showed in the following sections.

3.3 Identifying the linearly independent constraints

Not all the constraints obtained applying naïvely the separability definitions are necessary to define a Translog function as separable. But how many can be dropped? And which ones? Here we provide both a rule to calculate the number of independent constraints and a way to identify the ones of interest (i.e. those that are essential for describing the particular inputs partition chosen). The approach we propose is based on a comparison between nested CES functions and the Translog.

3.3.1 Theoretical tools

In the empirical literature concerning production, the most employed functions are the CES and the Translog: the former for its tractability given its convenient properties (i.e. homogeneity and separability) and the constancy of its elasticities, the latter for its flexibility and generality.

A first step in investigating the relationship between these two functional forms was made by Denny and Fuss (1977) who showed how a homogeneous and strongly separable Translog can approximate an unnested three-input CES when evaluated at a given point. However, this relationship becomes even more interesting when we consider the more recent class of nested CES. These nested functions are broadly employed to describe production technologies that are based on multiple production stages in which pairs or groups of inputs are separately combined to produce intermediate inputs. Thus, the way they are nested reflects a particular input separability structure. A comparison between them and the Translog can therefore throw light on the separability structure characterizing the production function under analysis.

A limitation of this approach, though, is given by the fact that nested CES and Translog functions are not directly comparable in their parameters as the former is non-linear while the latter is a linear logarithmic transformation of the inputs. To overcome this issue, we take a second order Taylor approximation in logarithms of the nested CES and exploit its linearisation. As it will be shown in the remainder of the chapter, the linearised CES has the functional form of a Translog but it is composed of a combination of the nested CES coefficients, thus allowing a very straightforward comparison.

3.3.1.1 Nested CES function

CES is a class of production functions characterized by the constancy of the elasticity of substitution which were originally investigated for the two-input case in the seminal paper by Arrow et al. (1961). Subsequently, numerous attempts were made to try to extend the CES concept to the n-input case. Two are the accepted extensions: the one-level n-input CES by Blackorby and Russell (1989) and the nested CES by Sato (1967).

Blackorby and Russell's (1989) extension is characterized by a single constant elasticity of substitution. Formally:[7]

$$Q = \lambda \left(\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right)^{-\frac{\nu}{\rho}}, \qquad \sum_{i=1}^{n} \delta_i = 1, \tag{3.10}$$

where $x = (x_1, \ldots, x_n)$ is the set of inputs, $\lambda > 0$ is the efficiency parameter $0 < \delta < 1$ is the share parameter, $\rho \in (-1, 0) \cup (0, \infty)$ is the substitution parameter and $\nu > 0$ is the scale parameter. The constant elasticity of substitution can be derived as $\sigma = 1/(1 + \rho)$.

Sato's (1967) describes a two-level n-inputs family of CES functions that allows for different nestings (i.e. subsets) of inputs, i.e.

$$Q = \lambda \left(\sum_{s=1}^{r} \delta_s (X_s)^{-\rho} \right)^{-\frac{\nu}{\rho}}, \quad \sum_{s=1}^{r} \delta_i = 1 \tag{3.11}$$

where

$$X_s = \lambda^{(s)} \left(\sum_{i=1}^{N_s} \delta_i^{(s)} (x_i)^{-\rho^{(s)}} \right)^{-\frac{\nu^{(s)}}{\rho^{(s)}}}, \quad \sum_{i=1}^{N_s} \delta_i^{(s)} = 1. \tag{3.12}$$

Equation (3.12) shows the inner level of the nested CES where the n inputs are combined in r subsets X_s that have a CES form. Equation (3.11) represents the outer level CES that combines the different subsets. When $\rho = \rho^{(s)}$ the two-level CES reduces to the plain one-level n-input CES.

[7]Note that this notation differ slightly from that of the other chapters as subset parameters are denoted with an exponent (s) instead of a subscript x. This simplifies the exposition given the large number of subsets.

Without loss of generality, the inner scale parameter $\nu^{(s)}$ can be imposed equal to one. Indeed, the outer scale parameter ν is already accounting for any change in the units of measure concerning the nested inputs. Also the inner efficiency parameter $\lambda^{(s)}$ must be normalised to one as, otherwise, one cannot separately identify the remaining CES parameters.[8]

3.3.1.2 Linearised CES properties

The linearisation of a CES function was illustrated by Kmenta (1967) and Hoff (2014) for the two-input and one-level n-input cases respectively. However, as we are studying separability, we are particularly interested in the nested case, which has hitherto not been linearised. Therefore, hereafter we outline an approach that can be followed to linearise a two-level three-input nested CES.[9] Following Kmenta (1967) and Hoff (2014), we use a Taylor expansion around the point where the substitution parameters equal one. With nested CES, however, the expansion is multivariate as we have more than one substitution parameter.

Let us consider a three-input two-level CES production function of the form $Q(x) = Q((x_1, x_2), x_3)$. The outer level is represented by

$$Q = \lambda\Big(\delta X_1^{-\rho} + (1-\delta)x_3^{-\rho}\Big)^{-\frac{\nu}{\rho}} \tag{3.13}$$

and the inner level is

$$X_1 = \gamma^{(1)}\Big(\delta^{(1)}x_1^{-\rho^{(1)}} + \left(1-\delta^{(1)}\right)x_2^{-\rho^{(1)}}\Big)^{-\frac{1}{\rho^{(1)}}}. \tag{3.14}$$

After substituting (3.14) into (3.13) and taking logarithms, we take a second order Taylor approximation in logarithms around $(\rho, \rho^{(1)}) = (0, 0)$. We first need to calculate the logarithm of (3.13) at $(0, 0)$:

$$f(0,0) = \ln(\gamma) + \delta\delta^{(1)}\nu\ln(x_1) + \delta\nu\left(1-\delta^{(1)}\right)\ln(x_2) + (1-\delta)\nu\ln(x_3) \tag{3.15}$$

[8] See van der Werf (2008), Baccianti (2013), and Henningsen and Henningsen (2012) for a discussion on this point.

[9] In Appendix A we present the linearisation of all the feasible three-input and four-input nested CES. Unfortunately, it is not possible to define a general rule for the linearisation of nested CES as, when the number of inputs increases, more and more nesting alternatives become available and each of them leads to a different linearisation.

Then, we need the gradient of the logarithm of (3.13) at (0,0):

$$
\begin{aligned}
\triangle_{1,1}(0,0) = {}& 0.5\delta\nu\left(\delta^{(1)}\right)^2(\delta-1)\ln^2(x_1)\\
&+0.5\delta\nu\left(\delta^{(1)}-1\right)^2(\delta-1)\ln^2(x_2)\\
&+0.5\delta\nu(\delta-1)\ln^2(x_3)\\
&-\delta\delta^{(1)}\nu(\delta-1)\left(\delta^{(1)}-1\right)\ln(x_1)\ln(x_2)\\
&-\delta\delta^{(1)}\nu(\delta-1)\ln(x_1)\ln(x_3)\\
&+\delta\nu(\delta^{(1)}-1)(\delta-1)\ln(x_2)\ln(x_3)\\
\triangle_{1,2}(0,0) = {}& 0.5\delta\delta^{(1)}\nu(\delta^{(1)}-1)\ln^2(x_1)+0.5\delta\delta^{(1)}\nu\left(\delta^{(1)}-1\right)\ln^2(x_2)\\
&-\delta\delta^{(1)}\nu\left(\delta^{(1)}-1\right)\ln(x_1)\ln(x_2)
\end{aligned}
\tag{3.16}
$$

The second order Taylor approximation of $\ln(Q)$, i.e. the linearised CES, is given by:

$$
\ln(\tilde{Q}) \cong ln(0,0) + \triangle_{1,1}(0,0)\rho + \triangle_{1,2}(0,0)\rho^{(1)},
$$

that is:

$$
\begin{aligned}
\ln(\tilde{Q}) \cong {}& \ln(\gamma)+\delta\delta^{(1)}\nu\ln(x_1)+\delta\nu\left(1-\delta^{(1)}\right)\ln(x_2)+(1-\delta)\nu\ln(x_3)+\\
&+0.5\delta\delta^{(1)}\nu\left(\delta^{(1)}\left(\rho(\delta-1)+\rho^{(1)}\right)-\rho^{(1)}\right)\ln^2(x_1)+\\
&+0.5\delta\nu(\delta^{(1)}-1)\left(\rho(\delta-1)\left(\delta^{(1)}-1\right)+\delta^{(1)}\rho^{(1)}\right)\ln^2(x_2)+\\
&+0.5(\delta-1)\delta\nu\rho\ln^2(x_3)+\\
&-\delta\delta^{(1)}\nu\left(\delta^{(1)}-1\right)\left(\rho(\delta-1)+\rho^{(1)}\right)\ln(x_1)\ln(x_2)+\\
&-\delta\delta^{(1)}\nu\rho(\delta-1)\ln(x_1)\ln(x_3)+\\
&+\delta\nu\rho(\delta-1)(\delta_x-1)\ln(x_2)\ln(x_3).
\end{aligned}
\tag{3.17}
$$

At this stage, it is critical to verify whether the linearised CES shares with the non-linear CES the properties of homogeneity and separability. Indeed, while the Translog can be considered itself a linearisation of a nested CES function that does not share the same properties, the linearised CES could have acquired some or all of its properties. Since, as previously anticipated, equation (3.17) can be seen as a Translog function where coefficients are combinations of CES parameters, we can use for the tests the constraints provided by Berndt and Christensen (1973b) for a general Translog as in equation (3.5). Homogeneity of degree v requires:

$$\begin{aligned} a_1 + a_2 + a_3 &= v \\ a_{11} + a_{12} + a_{13} &= 0 \\ a_{12} + a_{22} + a_{23} &= 0 \\ a_{13} + a_{23} + a_{33} &= 0. \end{aligned} \tag{3.18}$$

Whereas a Translog does not satisfy the constraints by construction, the linearised CES does: substituting the linearised CES coefficients of (3.17) in (3.18) it is straightforward to see that these are always true. Therefore, the linearised CES is a homogeneous function of degree v.

For what concerns separability, we need to jointly impose:

$$C1: \quad a_1 a_{23} - a_2 a_{13} = 0 \tag{3.19a}$$

$$\begin{aligned} C2: \quad & a_{11} a_{23} - a_{12} a_{13} = 0 \\ & a_{12} a_{23} - a_{22} a_{13} = 0. \end{aligned} \tag{3.19b}$$

Again, we need to substitute the coefficients of (3.17) in constraints $C1$ and $C2$ and verify if they are satisfied. It is trivial to see that albeit the $C1$ constraint is always satisfied, the $C2$ constraints hold only if

$$\rho = 0. \tag{3.20}$$

As the substitution parameter ρ cannot be null for construction, equation (3.20) is never satisfied. However, when ρ tends to zero we can say that the linearised CES is approximately separable. Indeed, this is the situation in which the approximation error is the smallest and the linearised CES best represents the underlying non-linear CES.[10]

For what concerns linear separability, this is given by the following constraints:

$$\begin{aligned} a_{23} &= 0 \\ a_{13} &= 0 \end{aligned} \tag{3.21}$$

or equivalently, using the linearised CES coefficients,

$$\begin{aligned} \delta\delta^{(1)} v\rho(\delta - 1) &= 0 \\ \delta v\rho(\delta - 1)\left(\delta^{(1)} - 1\right) &= 0. \end{aligned} \tag{3.22}$$

[10] This is in line with Danny et al.'s (1978) findings on approximate separability.

As ν must always be greater than zero and both δ and $\delta^{(1)}$ must lie between 0 and 1, these two constraints cannot be satisfied. Again, we need only ρ tending to zero to get closer to separability.

We conclude that we can use the linearised CES as a version of the Translog that embodies the constraints of homogeneity and separability of the class $C1$. Even if we do not explicitly discuss each case, we found this result to be true for any n-input linearised CES.

3.3.2 Number of independent constraints

Identifying the number of independent constraints can be a demanding task, especially for production technologies involving a large number of inputs. However, being able to reduce the number of constraints when testing for separability is of fundamental importance as it increases the degrees of freedom of the test and, thus, improves its preciseness.

For this purpose we provide a general rule that can be employed to find the number of independent constraints. The rule is based on the comparison between the Translog and the CES and the fact that the separability condition chosen can be seen in terms of CES nestings. Let us take as an example the case proposed in Section 3.2.1: we want to test whether three of the inputs can be separated from the fourth. This can be seen in terms of nesting as testing for a two-level four-input CES where the inner nest is composed by the three separable inputs.

The rule is the following:

$$N_{C1} = \frac{n!}{2!(n-2)!} - e \tag{3.23}$$

where N_{C1} represents the number of independent constraints of the type $C1$. The first term on the right hand side of (3.23) represents the number of easy combinations that can be obtained using pairs of inputs, i.e. the total number of elasticities that can be found given n inputs. In the example presented this term is 6. The second term e is the number of constant and different elasticities that characterize the corresponding nested CES. In this example, e is equal to two. Hence, N_{C1} represents the number constraints we need to impose on the remaining elasticities.

In order to find N_{C2}, i.e. the number of constraints of type $C2$, a general rule would be $N_{C2}(n-1)$. However, we would still have repeated constraints or constraints that are linear combination of some of the others. It is possible to analytically simplify the system of N_{C2} non-linear constraints to find the one those that are linearly independent. We report them in Table 3.1.

	N_{c1}	N_{c2}	$C1$	$C2$
(x_1, x_2, x_3)	2	3	$a_1a_{23} - a_2a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_1a_{23} - a_3a_{12} = 0$	$a_{13}^2 - a_{11}a_{33} = 0$
				$a_{11}a_{23} - a_{12}a_{13} = 0$
$((x_1, x_2), x_3)$	1	2	$a_1a_{23} - a_2a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
				$a_{11}a_{23} - a_{12}a_{13} = 0$
(x_1, x_2, x_3, x_4)	5	6	$a_1a_{23} - a_2a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_1a_{23} - a_3a_{12} = 0$	$a_{13}^2 - a_{11}a_{33} = 0$
			$a_2a_{34} - a_3a_{24} = 0$	$a_{14}^2 - a_{11}a_{44} = 0$
			$a_2a_{34} - a_4a_{23} = 0$	$a_{11}a_{23} - a_{12}a_{13} = 0$
			$a_3a_{24} - a_4a_{23} = 0$	$a_{11}a_{24} - a_{12}a_{14} = 0$
				$a_{11}a_{34} - a_{13}a_{14} = 0$
$((x_1, x_2)(x_3, x_4))$	3	5	$a_1a_{23} - a_2a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_3a_{14} - a_4a_{13} = 0$	$a_{11}a_{23} - a_{12}a_{13} = 0$
			$a_3a_{24} - a_4a_{23} = 0$	$a_{11}a_{24} - a_{12}a_{14} = 0$
				$a_{33}a_{14} - a_{13}a_{34} = 0$
				$a_{44}a_{13} - a_{14}a_{34} = 0$
$((x_1, x_2, x_3), x_4)$	4	5	$a_1a_{13} - a_3a_{12} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_2a_{13} - a_3a_{12} = 0$	$a_{13}^2 - a_{11}a_{33} = 0$
			$a_2a_{34} - a_3a_{24} = 0$	$a_{11}a_{23} - a_{12}a_{13} = 0$
			$a_1a_{24} - a_2a_{14} = 0$	$a_{11}a_{24} - a_{12}a_{14} = 0$
				$a_{11}a_{34} - a_{13}a_{14} = 0$
$((x_1, x_2), x_3, x_4)$	4	6	$a_1a_{34} - a_4a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_1a_{34} - a_3a_{14} = 0$	$a_{13}^2 - a_{11}a_{33} = 0$
			$a_2a_{34} - a_4a_{23} = 0$	$a_{14}^2 - a_{11}a_{44} = 0$
			$a_2a_{34} - a_3a_{24} = 0$	$a_{11}a_{23} - a_{12}a_{13} = 0$
				$a_{11}a_{24} - a_{12}a_{14} = 0$
				$a_{11}a_{34} - a_{13}a_{14} = 0$
$(((x_1, x_2), x_3), x_4)$	3	5	$a_1a_{23} - a_2a_{13} = 0$	$a_{12}^2 - a_{11}a_{22} = 0$
			$a_1a_{34} - a_3a_{14} = 0$	$a_{13}^2 - a_{11}a_{33} = 0$
			$a_2a_{34} - a_3a_{24} = 0$	$a_{11}a_{23} - a_{12}a_{13} = 0$
				$a_{11}a_{24} - a_{12}a_{14} = 0$
				$a_{11}a_{34} - a_{13}a_{14} = 0$

TABLE 3.1: Translog separability constraints in the three-input and four-input cases

3.3.3 Identifying the necessary constraints

The last step in order to identify which separability constraints are necessary and sufficient is comparing the linearised CES and the Translog coefficients. To this purpose, we need to write a system of $(3n + 1)$ identities and solve them for the Translog coefficients.[11] As the

[11]The quicker resolution approach is first to find the CES coefficients in terms of the Translog ones and then substitute them back into the original system.

number of Translog coefficients is larger than the number of CES coefficients, we find a number of constraints that is equal to the difference between the two.[12]

Once solved for the Translog coefficients, the constraints we obtain are $(n+1)$ homogeneity constraints and the N_{C1} linearly independent separability constraints we were looking for.

Although this method is applicable to any n-input case, in Table 3.1 we provide explicit solutions for all the feasible three-input and four-input cases. [13]

3.3.4 Consequences of the assumption of linear homogeneity

Previous literature has often assumed input homogeneity of degree ν to describe the production function returns to scale.[14] Imposing linear homogeneity further reduces the number of constraints required for $C2$ as showed by Berndt and Christensen (1973b). The authors provided an example of a three-input Translog function $F(x) = F(x_1, x_2, x_3)$ where they wanted to test for the separability structure $F(x) = F(G(x_1, x_2), x_3)$. In this case, the number of $C1$ constraints was limited to one and the number of $C2$ constraints to two. The authors showed that, when assuming constant returns to scale, it is possible to rewrite these conditions using only five of the nine Translog parameters as follows:

$$C1: \quad a_3 = 1 + (a_2 a_{23}/a_{22}) \tag{3.24a}$$

$$C2: \quad a_{23}^2 - a_{22}a_{33} = 0. \tag{3.24b}$$

However, it can be shown analytically that when assuming homogeneity of degree ν, the number and the expression of $C2$ conditions is the same for each separability structure and vary only with the number of inputs considered.[15] While for the three-input case the constraint is given by (3.24b), the four-input case requires the following set of $C2$

[12]Indeed, another method to determine the number of constraints required by the Translog is given by the difference between the number of CES and Translog coefficients minus the number of constraints required by homogeneity (i.e. $n + 1$).

[13]The resolution method for the system of equations is not unique. Thus, the expressions for the $C1$ constraints can vary. Nevertheless, they are all equivalent.

[14]The production function is characterized by constant returns to scale (i.e. it is linearly homogeneous) when $\nu = 1$, decreasing when $\nu < 1$ and increasing when $\nu > 1$.

[15]In order to attain the reported constraints one needs to take the $C2$ constraints as in Table 3.1 and substitute in each of them the homogeneity conditions $a = -(a_{22} + a_{23} + a_{24})$, $a_{13} = -(a_{23} + a_{33} + a_{34})$, $a_{44} = -(a_{24} + a_{34} + a_{44})$, $a_{11} = a_{22} + a_{33} + a_{44} + 2a_{12} + 2a_{13} + 2a_{14} + 2a_{23} + 2a_{24} + 2a_{34}$ and look at which constraints are repeated or are a combination of the others.

constraints to be satisfied:

$$
\begin{aligned}
a_{23}^2 - a_{22}a_{33} &= 0 \\
a_{34}^2 - a_{33}a_{44} &= 0 \\
a_{33}a_{24} - a_{23}a_{34} &= 0.
\end{aligned}
\tag{3.25}
$$

To find the corresponding $C1$ constraints it is enough to substitute each a_1 term with $(\nu - \sum_{i=2}^{n} a_i)$ and simplify.

3.4 Conclusions

The existing empirical literature has often taken advantage of the assumption of input separability, however it has very rarely tested it. The few theoretical and applied works which defined separability for a general production function and studied the relative constraints for the Translog in particular, have mainly focused on three-input cases or simple separability structures leaving the reader the task of formalizing the conditions for other more complex cases. However, it is not straightforward to identify which one are the appropriate separability constraints to impose on the estimated coefficient of a Translog production function with more than three inputs.

In this chapter, we have shown an approach that helps identify the number and the type of constraints that are necessary and sufficient to test separability. This is based on the linearisation of nested CES functions, whose algebraic resolution is presented for the first time. While we explicitly provide these constraints for the three-input and four-input cases, the procedure is general and can be employed with any n-input Translog.

Chapter 4

Is the Production Function CES?

An empirical procedure to help discriminate between functional forms

4.1 Introduction

The empirical literature on the estimation of the substitution relationships between energy and other inputs has been growing since the burst of the oil crisis in 1973. Since then, the econometric methods have evolved as well as the scopes of studies: while first papers were driven by productivity concerns following the oil crisis, more recent papers are interested in assessing the impact of climate change and environmental policies on production. The vast majority of the papers has exploited a Translog production function as it can be easily adopted in diverse application contexts. Indeed, this functional form is general, in the sense that it allows to test different assumptions on inputs and technology rather than maintaining them, it is log-linear and thus easy to estimate even in specifications where some of the classical regression assumptions are violated, and it is analytically convenient for deriving factor demand functions and the cost function. Nevertheless, a small fraction of the most recent work favoured constant elasticities of substitution (CES) functions. Many of these papers belong to the Computable General Equilibrium (CGE) literature, which lately has recognized the importance of empirically informed parameters for its models. This strand of literature has traditionally taken advantage of the convenient maintained properties of CES (and the special cases it nests i.e. Cobb-Douglas and Leontief) as they guarantee the function to be globally well-behaved and tractable. These characteristics are particularly advantageous in a CGE framework as they simplify the model computationally and help ensure the convergence of its numerical solution.

In addition to the estimation problems deriving from its non-linear form and the fact that production data are usually short time-series characterized by serial and simultaneous correlation, there are two main issues connected with the use of a CES production function in empirical applications. Firstly, this functional form is highly restrictive as it is based on maintained hypotheses that are often not consistent with empirical applications.[1] Secondly, when using more than two inputs, nested CES functions allow greater flexibility in terms of substitution relationships between inputs, but this implies that researchers are compelled to define a nested structure for the inputs. Although certain structures which do not make economic sense could be disregarded *a priori*,[2] this choice should be motivated and supported by formal selection procedures.

Kemfert (1998) suggested the use of the R^2 statistic to discriminate between nested structures in the three-input case: among the $((E,K),L)$, $((E,L),K)$, and $((K,L),E)$ alternatives, the one with the highest statistic is selected. This method was then used by van der Werf (2008) and Baccianti (2013). There are some drawbacks connected with this selection procedure. First, model selection criteria should only be used when the applied researcher is convinced that the set of models considered includes the true model. However, in this instance we cannot exclude *a priori* the possibility that the true underlying functional form is not consistent with a nested CES and that there is another functional form that would provide a better representation of the true input-output relationship. If that would be the case, one of the nested structures would always be favoured even if none represents the "best" characterization of true production function. Second, Kemfert (1998) as well as the subsequent authors, did not consider the unnested case, (E,K,L), among the feasible structures. Since this is characterized by the same number of variables, but by a smaller number of parameters, the adjusted R^2 should be preferred. Thirdly, the R^2 statistic cannot be used when the alternative nested CES structures are estimated using a system of conditional factor demands as the resulting econometric models have different dependent variables.

In this chapter, we propose and explore a new empirical procedure that tackles at once both issues connected with the use of a CES production function. In particular, it can be used to both understand whether for a given dataset the unknown production function is consistent with a CES, and to discriminate between alternative nested structures. It also provides a link between the applied econometrics and CGE literature as it rests upon a Translog functional form whose coefficients can be tested for some of the CES maintained hypotheses (i.e. homogeneity and separability). The reason we chose a Translog among other general functional forms is that the connection with the CES is straightforward: when the Translog coefficients satisfy specific constraints implied by the CES, it can be interpreted as a second order Taylor approximation to an arbitrary CES. We consider both

[1] See, among others, Hazilla and Kopp (1986), Iqbal (1986), Garofalo and Malhotra (1988), Khiabani and Hasani (2010), and Haller and Hyland (2014) for application in which the Translog homogeneity or separability conditions were rejected.

[2] E.g. a structure with an intermediate input formed by labour and energy such as $((E,L),K)$ where E is energy, L is labour, and K is capital.

the two-input and the three-input cases, where the first is used as a baseline to understand if the procedure performs correctly even in the absence of separability assumptions.

The suggested procedure is the following. The first phase consists in a number of tests performed on the Translog coefficients in order to understand if the estimated Translog is consistent with a linearised CES. In particular, we use a Monte Carlo simulation framework, where the data generating process is based on a CES production function, to compare different inference tests and evaluate which one performs best in terms of size and power within different parametrisations. A failure to reject the tested restrictions represents a first indication that a CES could be the appropriate function to describe the input-output relationship. Moreover, with more than two inputs, the test also informs on which nested CES more closely approximates the true one. Thus, our approach to the selection between nested structures is not based on comparisons of goodness of fit measures, but it has a theoretical foundation and acknowledges the possibility that the true production function might not be consistent with a CES.

In empirical applications, the results of the first phase can deliver two outcomes. On the one hand, the results may indicate that we reject some of the maintained characteristics of the non-linear CES and, thus, the procedure concludes that a CES is not appropriate for that specific dataset. On the other hand, results may be consistent with a CES model, and the second phase of the procedure is used to understand if the underlying model is a non-linear CES or just is better view as its approximation.

The second phase consists in both a graphical analysis and formal selection tests. We derive the point substitution elasticities of the linearised CES and prediction intervals around them. If we observe peaks in their distribution around a small range of values and narrow prediction intervals, we can conclude that the dataset supports the hypothesis of a constant elasticity (i.e. a CES structure is appropriate). The formal tests consist in computing different selection criteria to determine which of the two rival models performs better.

The rest of the chapter is organised as follows: Section 4.2 describes the Monte Carlo approach that is used throughout the chapter and provides an empirical explanation behind the approximation and estimation errors. Section 4.3 and Section 4.4 outline the first and the second phases respectively for the two and three inputs cases. Finally, Section 4.5 concludes.

4.2 Monte Carlo simulation approach

In the following sections, we run a number of Monte Carlo simulations[3] with the aim of understanding how inference tests, used to identify the "true" underlying input-output relationship, perform in terms of size and power.

For simplicity, we consider only three inputs: energy (E), capital (K), and labour (L). The method outlined can be easily generalized, and for ease of exposition, here we only focus on the following two-input CES and three-input nested CES:

$$q_t^{CES} = f(E_t, K_t) = \ln(\lambda) - \frac{\nu}{\rho} \ln\left(\delta E_t^{-\rho} + (1-\delta) K_t^{-\rho}\right) \tag{4.1a}$$

$$\begin{aligned} q_t^{CES} =& f(E_t, K_t, L_t) = \ln(\lambda) - \frac{\nu}{\rho} \ln\Big(\delta \left(\delta_x E_t^{-\rho_x} + (1-\delta_x) K_t^{-\rho_x}\right)^{\rho/\rho_x} + \\ & + (1-\delta) L_t^{-\rho}\Big) \end{aligned} \tag{4.1b}$$

where $\lambda > 0$ is the efficiency parameter, $\rho, \rho_x \in (-1, 0) \cup (0, \infty)$ are the substitution parameters,[4] $\nu > 0$ is the scale parameter, $\delta, \delta_x \in (0, 1)$ are the share parameters, and $t = 1, \ldots, T$ indexes observations. The constant elasticities of substitution can be derived as $\sigma = 1/(1+\rho)$ and $\sigma_x = 1/(1+\rho_x)$. Note that when $\rho = 0$, the CES reduces to a Cobb-Douglas; when $\rho < 0$, $\sigma > 1$; when $\lim_{\rho\to\infty}$ the production function approaches a Leontief.

We define two Data Generating Processes (DGPs), one for the two-input case (DGP1) and one for the nested 3-input case (DGP2). In both of them, output is generated according to the following specification:

$$y_t = q_t^{CES} + \epsilon_t \tag{4.2}$$

where y_t is the logarithm of output Y_t, and ϵ_t is a normally distributed error term with mean equal to zero and variance equal to σ_ϵ. The values of the parameters and the distributions of the inputs are listed in Table 4.1.

In both DGPs, we let the substitution parameter(s) and the variance of the disturbance term vary across certain ranges of values. Indeed, both parameters greatly influence the CES estimates: the substitution parameter affects the overall curvature of the CES and, hence, the ease with which it can be fit; the variance of the disturbances influences the deviation

[3] Unless otherwise specified, the number of Monte Carlo simulations is 1000 in each case. In this chapter, we have used Stata 13 by StataCorp (2013) and the following user written program: Mander (2005).

[4] The subscript x indicates that the parameter refers to the inner nests.

	DGP1	DGP2
E	$\sim \ln N(0, 0.5)$	$\sim \ln N(0, 0.5)$
K	$\sim \ln N(0, 0.5)$	$\sim \ln N(0, 0.5)$
L	-	$\sim \ln N(0, 0.5)$
δ	0.5	0.5
δ_x	-	0.5
ν	1	1
T	1,000	1,000

TABLE 4.1: Data Generating Processes

of the output observations from the CES output values.[5] In DGP1, ρ assumes eight values as shown in Table 4.2 and the variance of the disturbances assumes four values, i.e. 0.01, 0.05, 0.1, and 0.5. In DGP2, since we have two substitution parameters, we slim down our analysis and limit ρ and ρ_x to six values (i.e. we exclude the -0.4 and 0.4 cases). We arbitrarly selected these levels as they include a range of values that spans from very low to very high substitutability for the substitution parameter and very low to very high variance for the disturbances. We tried alternative parameterizations with wider ranges and smaller intervals between the levels but we believe these values to be the most informative for the aim of the chapter.

ρ	-0.9	-0.4	-0.1	0.1	0.4	0.9	2	9
σ	10	1.67	1.11	0.91	0.71	0.53	0.33	0.1

TABLE 4.2: Selected values for the substitution parameter and the corresponding elasticities of substitution

Finally, for each test, we repeat the simulations altering the input distributions and the remaining CES parameters one at a time in order to evaluate how results are affected.

4.2.1 Measure of the bias of the Translog model

When one uses a log-linear model to estimate a non-linear relationship, she incurs in a model bias. In particular, in the Translog case, the bias is explained by the fact that the coefficients of the Translog are unable to capture interactions between inputs and output of order higher than two. In this section, we exploit Monte Carlo simulations to obtain a measure of this bias, and how it is affected by a changes in the values of the substitution parameters.

[5] These concepts will be further discussed in the next section.

As the substitution parameters increase, the CES becomes more curved and, thus, more difficult to estimate using a log-linear model. In Figure 4.1, we show, for the two-input case, the CES assumed in DGP1 (coloured and reticulated surface) and the estimated CES obtained from the Translog regression (grey and plain surface) when the variance of the disturbances is imposed to be null. In this case, the total distance between the two surfaces (i.e. the sum of residuals from the estimation in absolute terms) represents the bias that occurs from using a second order log-linear model to estimate a CES model, and this becomes bigger as ρ increases.

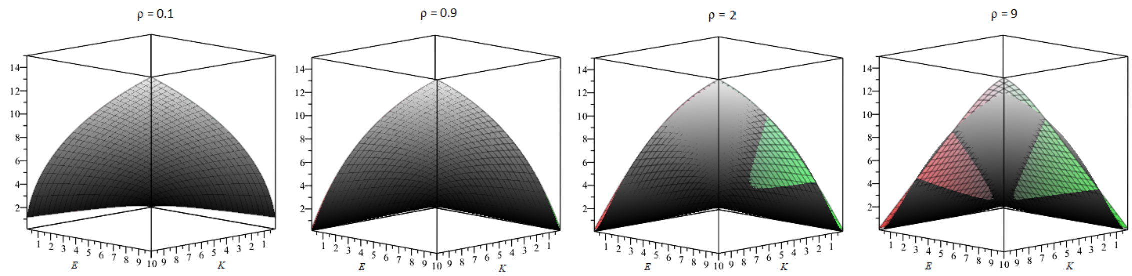

FIGURE 4.1: Bias from the Translog estimation in the two-input case for different values of the substitution parameter

Suppose the two-input and three-input Translog are given by, respectively[6]

$$\begin{aligned} q_t^{UT} = g(E_t, K_t) =& a_0 + a_1 \ln(E_t) + a_2 \ln(K_t) + \\ & + 0.5 a_{11} \ln^2(E_t) + 0.5 a_{22} \ln^2(K_t) + a_{12} \ln(E_t) \ln(K_t) \end{aligned} \tag{4.3a}$$

$$\begin{aligned} q_t^{UT} = g(E_t, K_t, L_t) =& a_0 + a_1 \ln(E_t) + a_2 \ln(K_t) + a_3 \ln(L_t) + \\ & + 0.5 a_{11} \ln^2(E_t) + 0.5 a_{22} \ln^2(K_t) + 0.5 a_{33} \ln^2(L_t) + \\ & + a_{12} \ln(E_t) \ln(K_t) + a_{13} \ln(E_t) \ln(L_t) + a_{23} \ln(K_t) \ln(L_t). \end{aligned} \tag{4.3b}$$

Algebraically, we define the mean squared bias as:

$$MSB = \frac{1}{N} \sum \left(q_t^{CES} - \hat{q}_t^{UT} \right)^2 \tag{4.4}$$

where $\hat{q}_t^{UT}$ are the fitted values from the OLS estimation of y_t using a Translog as in (4.3a) or (4.3b) with an added error term.

Table 4.3 and 4.4 report the mean squared bias for different values of ρ and σ_ϵ for DGP1 and DGP2[7] respectively. As predicted, the bias increases with the substitution parameter.

[6]UT is mnemonic for unconstrained Translog.

[7]We only consider positive values for the substitution parameters as results are approximately symmetric around zero in both directions.

Moreover, the bias is negligible, when ρ is smaller than one in absolute terms.

$\sigma_\epsilon \setminus \rho$	-0.9	-0.4	-0.1	0.1	0.4	0.9	2	9
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.05	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003	0.0038
0.1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003	0.0038
0.5	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0018	0.0053

TABLE 4.3: Mean squared bias for DGP1

$\rho_x \setminus \rho$	0.1	0.9	2	9	0.1	0.9	2	9
	$\sigma_\epsilon = 0$				$\sigma_\epsilon = 0.01$			
0.1	0.0000	0.0000	0.0001	0.0026	0.0000	0.0000	0.0001	0.0026
0.9	0.0000	0.0001	0.0004	0.0033	0.0000	0.0001	0.0004	0.0033
2	0.0001	0.0003	0.0010	0.0048	0.0001	0.0003	0.0010	0.0048
9	0.0009	0.0015	0.0030	0.0084	0.0009	0.0015	0.0030	0.0084
	$\sigma_\epsilon = 0.05$				$\sigma_\epsilon = 0.1$			
0.1	0.0000	0.0000	0.0002	0.0026	0.0001	0.0001	0.0002	0.0027
0.9	0.0000	0.0001	0.0004	0.0034	0.0001	0.0002	0.0005	0.0034
2	0.0001	0.0003	0.0010	0.0048	0.0002	0.0004	0.0011	0.0049
9	0.0010	0.0015	0.0030	0.0084	0.0010	0.0016	0.0031	0.0085
	$\sigma_\epsilon = 0.5$							
0.1	0.0025	0.0025	0.0026	0.0051				
0.9	0.0025	0.0026	0.0029	0.0058				
2	0.0025	0.0028	0.0035	0.0073				
9	0.0034	0.0040	0.0055	0.0109				

TABLE 4.4: Mean squared bias for DGP2

We also observe that the bias increases with the variance of the disturbances, and that each increase is approximately constant for different ρ. Indeed, the higher the variance of the disturbances, the more the estimated model describes the noise instead of the underlying CES.

The only difference between the two DGPs is given by the magnitude of the MSBs: in DGP2 they are overall larger and grow faster with the substitution parameters. In particular, in the three-input case, results are more affected by changes in ρ_x than ρ.

The relationship between the residuals form the Translog estimation ($\hat{\epsilon}_t^{UT}$) and the bias is made explicit in the following expressions:

$$\begin{aligned} \hat{\epsilon}_t^{UT} &= y_t - \hat{q}_t^{UT} \\ &= \epsilon_t + (q_t^{CES} - \hat{q}_t^{UT}) \end{aligned} \tag{4.5}$$

where $\hat{q}_t^{UT}$ are the fitted values from the UT estimation and the term in the brackets is the bias. There is a positive relationship between residuals and both bias and the true disturbance term. Table 4.5 and Table 4.6 show the Mean Squared Error (MSE) of the Translog regression for DGP1 and DGP2 respectively. We can see that the impact of the bias is particularly pronounced when the variance of the disturbances is small.[8]

$\sigma_\epsilon \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0004	0.0039
0.05	0.0025	0.0025	0.0025	0.0025	0.0025	0.0025	0.0028	0.0063
0.1	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0103	0.0138
0.5	0.2498	0.2498	0.2498	0.2498	0.2498	0.2498	0.2501	0.2536

TABLE 4.5: Mean Squared Error for DGP1

$\rho_x \diagdown \rho$	*0.1*	*0.9*	*2*	*9*	*0.1*	*0.9*	*2*	*9*
	$\sigma = 0.01$				$\sigma = 0.05$			
0.1	0.0001	0.0001	0.0002	0.0027	0.0025	0.0025	0.0026	0.0051
0.9	0.0001	0.0002	0.0005	0.0035	0.0025	0.0026	0.0029	0.0051
2	0.0002	0.0004	0.0011	0.0050	0.0026	0.0028	0.0035	0.0051
9	0.0011	0.0016	0.0031	0.0086	0.0034	0.0040	0.0055	0.0051
	$\sigma = 0.1$				$\sigma = 0.5$			
0.1	0.0100	0.0100	0.0101	0.0126	0.2495	0.2495	0.2496	0.2521
0.9	0.0100	0.0101	0.0104	0.0134	0.2495	0.2496	0.2499	0.2529
2	0.0100	0.0103	0.0110	0.0149	0.2496	0.2498	0.2506	0.2544
9	0.0109	0.0115	0.0130	0.0185	0.2504	0.2510	0.2525	0.2580

TABLE 4.6: Mean Squared Error for DGP2

Finally, we can look at the described effects on the estimated CES parameters and the relative standard errors.[9] As the number of parameters is bigger in a Translog than in a CES function, there is not a single way of writing the CES parameters in terms of the Translog ones. However, in Table 4.7, we report the values for one of the possible combinations available: the purpose is only to observe how the magnitude of their bias and preciseness varies with the substitution parameter (and with σ_ϵ). We consider only the two-input case as findings for the three-input one are concordant. Table 4.7 confirms that the bias increases with ρ and σ_ϵ and that the standard error are not only affected by an increase in the variance of the disturbances but also by the substitution parameter. Moreover, we observe that $\hat{\lambda}$ and δ_x estimates are unbiased and precise across all parametrisations, $\hat{\delta}$ is slightly underestimated for high values of ρ and $\hat{\rho}$ is the most sensible to changes in the DGP parameters.

[8]In fact, when σ_ϵ increases, the value of the product between the error and the bias becomes larger in absolute terms.

[9]We consider only positive ρ as results are approximately symmetrical.

$\sigma_\epsilon \backslash \rho$	*0.1*	*0.4*	*0.9*	*2*	*9*	*0.1*	*0.4*	*0.9*	*2*	*9*
			$\hat{\lambda}$					$\hat{\nu}$		
0.01	1.500	1.500	1.497	1.480	1.385	1.000	1.000	1.000	1.000	1.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
0.05	1.500	1.500	1.497	1.480	1.385	1.000	1.000	1.000	1.000	1.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
0.1	1.500	1.499	1.496	1.479	1.385	1.000	1.000	1.000	1.000	1.000
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
0.5	1.498	1.498	1.495	1.478	1.383	1.000	1.000	1.000	1.000	1.001
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
			$\hat{\delta}$					$\hat{\rho}$		
0.01	0.500	0.500	0.500	0.500	0.500	0.100	0.392	0.825	1.460	2.190
	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.005)	(0.005)	(0.005)	(0.010)	(0.034)
0.05	0.500	0.500	0.500	0.500	0.500	0.099	0.392	0.824	1.459	2.190
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.015)	(0.015)	(0.016)	(0.018)	(0.038)
0.1	0.500	0.500	0.500	0.500	0.500	0.098	0.390	0.824	1.457	2.186
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.051)	(0.051)	(0.052)	(0.054)	(0.064)
0.5	0.500	0.500	0.500	0.501	0.502	0.090	0.380	0.819	1.447	2.182
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.256)	(0.256)	(0.259)	(0.264)	(0.276)

TABLE 4.7: Estimated CES parameters and standard errors (in parenthesis) from a Translog regression

4.2.2 Test on regularity conditions

Unlike the CES case, Translog functions are not characterized by global validity, in the sense that they are not always well-behaved, i.e. output increasing monotonically and convex isoquants. It is interesting to see whether the regularity conditions are satisfied for the DGPs we have specified. Results for DGP1 are summarized in Table 4.8. The first four rows refer to monotonicity and the latter four to convexity: we can see that, although the functional form assumed in the DGP is CES, the percentages in which the conditions are satisfied decrease at the extremes.

$\sigma_\epsilon \backslash \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	100	100	100	100	100	100	95	80
0.05	100	100	100	100	100	100	95	80
0.1	100	100	100	100	100	100	95	80
0.5	99	100	100	100	100	100	94	80
0.01	85	100	100	100	100	100	100	98
0.05	84	100	100	100	100	100	100	98
0.1	83	100	100	100	100	100	100	98
0.5	62	99	100	100	100	100	100	98

TABLE 4.8: Percentages of times the Translog satisfies monotonicity and convexity in DGP1

This is even more pronounced in the three-input case (Table 4.9 and 4.10) where we clearly observe that also an increase in the variance of the disturbance term have a negative impact. We can, thus, conclude that a Translog acquires global validity only for values of the substitution parameters close to zero, i.e. where the model bias is smaller.

	-0.9	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
	$\sigma_\epsilon = 0.01$						$\sigma_\epsilon = 0.05$					
-0.9	99	100	100	99	92	75	99	100	100	99	92	75
-0.1	100	100	100	100	96	80	100	100	100	100	96	80
0.1	100	100	100	100	96	80	100	100	100	100	96	80
0.9	99	100	100	99	91	73	99	100	100	99	91	73
2	93	95	95	90	81	64	93	95	95	90	81	64
9	81	81	80	76	68	53	81	81	80	76	68	53
	$\sigma_\epsilon = 0.1$						$\sigma_\epsilon = 0.5$					
-0.9	99	100	100	99	92	75	95	98	98	95	88	72
-0.1	100	100	100	100	96	79	98	100	100	98	91	75
0.1	100	100	100	100	96	79	98	100	100	98	91	75
0.9	99	100	100	99	91	73	96	98	98	95	87	71
2	93	95	94	90	81	64	89	91	91	87	78	62
9	81	81	80	75	68	53	78	78	78	74	66	52

TABLE 4.9: Percentages of times the Translog satisfies monotonicity in DGP2

	-0.9	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
	$\sigma_\epsilon = 0.01$						$\sigma_\epsilon = 0.05$					
-0.9	65	83	86	70	64	60	63	81	84	70	64	60
-0.1	88	100	100	100	97	87	87	100	100	100	97	87
0.1	88	100	100	100	99	92	87	100	100	100	99	92
0.9	82	100	100	100	100	97	81	100	100	100	100	97
2	73	100	100	100	99	97	72	100	100	100	99	97
9	63	98	98	98	98	95	63	98	98	98	98	95
	$\sigma_\epsilon = 0.1$						$\sigma_\epsilon = 0.5$					
-0.9	62	78	80	69	63	60	52	50	52	58	58	57
-0.1	83	100	100	100	97	87	49	94	95	94	90	82
0.1	84	100	100	100	99	92	52	96	97	97	94	87
0.9	78	100	100	100	100	97	54	98	99	100	99	95
2	70	100	100	100	99	97	52	97	98	99	98	95
9	61	98	98	98	98	95	50	94	96	97	96	93

TABLE 4.10: Percentages of times the Translog satisfies convexity in DGP1

4.3 First phase: hypothesis testing

Inference tests represent a first step towards understanding whether the true production function underlying a given dataset has a CES form. Indeed, CES are homogeneous and separable by construction, whereas Translog functions can be tested, at least in part, for these maintained assumptions. As shown in Chapter 2, the CES shares with its linearised version the maintained hypotheses of homogeneity and approximate separability. Hence, if we test the Translog coefficients for these assumptions, we are at once checking for evidence in favour of a CES and verifying whether the Translog is a linearised CES. The null hypothesis of the various testing techniques that we present in this section is that the estimated Translog parameters satisfy the homogeneity and separability (when we consider more than two inputs) restrictions.

Hypothesis testing not only provides evidence in favour or against a CES representation of the data, but also, when there are more than two inputs, informs on which nested CES is the most appropriate. From the joint test of homogeneity and approximate separability, we can find which, if any, nested structures are not rejected empirically: we can find that all, none, or only some of them are not rejected.

An unusual feature of our testing approach is that, although in the DGPs the true production function is CES, the null hypothesis is a linearised CES (hereafter CT, for constrained Translog). The implication is that results depend on the bias of the model: as we move away from the approximation point (i.e. when the substitution parameters become larger in absolute terms) the CES becomes more curved and the bias from the model larger, leading to the rejection of the restrictions on the coefficients. To prove this point, we include an additional set of results where the DGPs are based on a CT production function and the dataset remains unchanged: results should be unaffected by changes in the substitution parameters as both models are log-linear. The feasible outcomes that we can obtain from tests based on a DGP that is CES or CT are summarized in Table 4.11.

	CES	
CT	F,F	F,R
		R,R

TABLE 4.11: Possible testing outcomes for DGPs based on CES or CT functional forms. F stands for fail to reject, and R for reject.

The (R,R) case clearly indicates that neither CT nor CES functional forms are appropriate and a more general form should be favoured. The (F,F) case suggests that both forms can be deemed appropriate and further investigation is needed. Finally, under (F,R), the test results based on a true CES are biased but those based on a true CT indicate that the restrictions are verified.[10] We should then conclude that both functional forms are appropriate and that additional testing is needed to discriminate between them. However, unfortunately, in real applications we are not aware of the true functional form of the production function and we only observe one set of results. Even though failing to reject the constraints always implies that the production function could be CES, their rejection could be due to the bias from the linearisation.

4.3.1 Wald test

The first hypothesis testing method that we analyse is the Wald test. A Wald-type test has two main advantages, but also important shortcomings. The advantages are that it only requires the estimation of the unconstrained model, reducing computational burden, and that it does not rely on the assumption of normally distributed disturbances, allowing for the presence of heteroskedasticity and serial correlation. The two main shortcomings emerge only in the presence of non-linear constraints. The first one is that the outcome of the test may be biased by an error of approximation. In order to compute the variance of a non-linear combination of estimated parameters, the Wald test appeals to the Delta method: the non-linear restrictions are linearised using a first order Taylor approximation around the true parameter vector (Greene, 2008, p. 97). The second is that Wald-type tests are not invariant to different algebraic formulation of the hypotheses: when the null can be written in two alternative ways, the relative test results may lead to opposite conclusions (Gregory and Veall, 1985, Lafontaine and White, 1986).

In this section, we want to assess whether the Wald statistic correctly fails to reject the null of both linear homogeneity and non-linear separability when the assumed production function is based on a CES or a CT. Formally, let $H_0 : \boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{b}$ be the null hypothesis and $\hat{\boldsymbol{\theta}}$ be a vector of parameter estimates deriving from the unconstrained regression, in our case the Translog. The Wald statistic is:[11]

$$W = \left[\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right) - \boldsymbol{b}\right]' \left(AVar\left[\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right) - \boldsymbol{b}\right]\right)^{-1} \left[\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right) - \boldsymbol{b}\right]$$

[10] The (R,F) case is not examined as it is infeasible: the CT maintained conditions are all shared by a CES, but not *viceversa*.

[11] See Greene (2008, p. 501).

where $AVar$ stands for asymptotic variance. A large W leads to rejection of the null. In order to estimate the asymptotic variance, $E.AVar$, the Delta method is employed:

$$E.AVar\left[\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right)-\boldsymbol{b}\right]=\hat{\boldsymbol{C}}E.AVar\left[\hat{\boldsymbol{\theta}}\right]\hat{\boldsymbol{C}}'$$
$$\text{where}\qquad \hat{\boldsymbol{C}}=\frac{\partial\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right)}{\partial\hat{\boldsymbol{\theta}}'}.$$

When the restrictions are linear, $H_0:\boldsymbol{c}(\boldsymbol{\theta})=\boldsymbol{R}\boldsymbol{\theta}=\boldsymbol{b}$, the asymptotic variance reduces to:

$$E.AVar\left[\boldsymbol{c}\left(\hat{\boldsymbol{\theta}}\right)-\boldsymbol{b}\right]=\boldsymbol{R}E.AVar\left[\hat{\boldsymbol{\theta}}\right]\boldsymbol{R}'.$$

Thus, the Wald statistics depends positively on the square of $\boldsymbol{c}(\hat{\boldsymbol{\theta}})-\boldsymbol{b}$ and negatively on the asymptotic variance of the parameters.

4.3.1.1 Monte Carlo simulations with two inputs

Let us consider the benchmark DGP1. To test for linear homogeneity using a Wald test, we first estimate a Translog as in (4.3a) with an added disturbance term, and then test if the following three constraints are jointly satisfied by its estimated coefficients:

$$\begin{cases} a_1+a_2=1 \\ a_{11}+a_{12}=0 \\ a_{12}+a_{22}=0. \end{cases}\tag{4.6}$$

If we cannot reject the null, we conclude that the function is homogeneous of first degree.[12]

Table 4.12 summarizes the percentage of times we reject the null hypothesis in DGP1 for different values of ρ and σ_ϵ, i.e. the size of the test. As the functional form assumed in the DGP is CES and we choose a significance level of 5%, we expect to reject approximately 5% of the times in each scenario.

$\sigma_\epsilon \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	8.8	4.8	4.8	4.9	4.6	8.6	24.2	19.5
0.05	5.0	5.0	5.0	5.0	5.0	5.0	8.0	14.0
0.1	4.9	4.8	4.9	4.9	4.9	4.6	5.8	9.1
0.5	4.8	4.8	4.9	4.9	4.9	4.9	4.7	5.3

TABLE 4.12: Rejection levels for Wald tests on homogeneity (percentages) for DGP1

[12] We arbitrary set homogeneity to be of first degree as it coincides with constant returns to scale which is often assumed in empirical works. However, results are invariant to alternative homogeneity assumptions.

We observe that in most cases the rejection levels approximately equal 5%, as predicted. Moreover, as anticipated in the introduction of this section, the size of the test depends on the values of ρ: the further we move from the approximation point (i.e. the larger is ρ in absolute terms), the bigger the bias, the more we reject. This can be clearly seen in the first two rows of the table: the size of the test increases for values of ρ greater than unity. However, we also observe a small decrease when $\sigma_\epsilon = 0.01$ and $\rho = 9$ and the reason is that the standard errors (and thus the denominator of the Wald statistic) increase more in proportion than the bias of the coefficients. Furthermore, Table 4.12 shows that the size of the test clearly depends on σ_ϵ in a negative way: when σ_ϵ is large, standard errors increase and the Wald statistic decreases. This effect is accentuated by the fact that the model bias increases with the variance of the disturbances. Nevertheless, the size of the test in all the cases considered is such that we fail to reject the null at least 75% of the times.

We provide further evidence of the effect of the bias due to the linearisation in Table 4.13, where we present the results of the Monte Carlo simulations based on a DGP where the production function is CT: as expected, rejection levels remain approximately constant across all ρ specifications.

$\sigma_\epsilon \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	4.8	4.8	4.8	4.9	4.6	4.6	4.6	4.6
0.05	4.7	4.8	4.9	4.9	4.9	4.9	4.9	4.9
0.1	4.9	4.8	4.9	4.9	4.9	4.9	4.9	4.9
0.5	4.8	4.8	4.9	4.9	4.9	4.9	4.9	4.9

TABLE 4.13: Size of Wald tests on homogeneity (percentages) when DGP is CT

We now investigate the power of the test, i.e. how many times the test rejects a null that is false, gradually increasing and decreasing the right hand side of the three restrictions in (4.6) by 0.01 and looking at the rejection level in each case. We repeat the same procedure for different values of ρ and σ_ϵ. Since results do not vary with ρ, Figure 4.2 shows the four power curves corresponding to different values of σ_ϵ. What emerges is that an increase in the error variance reduces the power of the test. The reason is that the greater the error variance, the smaller the part of total output variation explained by the deterministic model. Nevertheless, all the power curves increase back to the 100% rejection level very rapidly and this suggests that the test has statistical power.

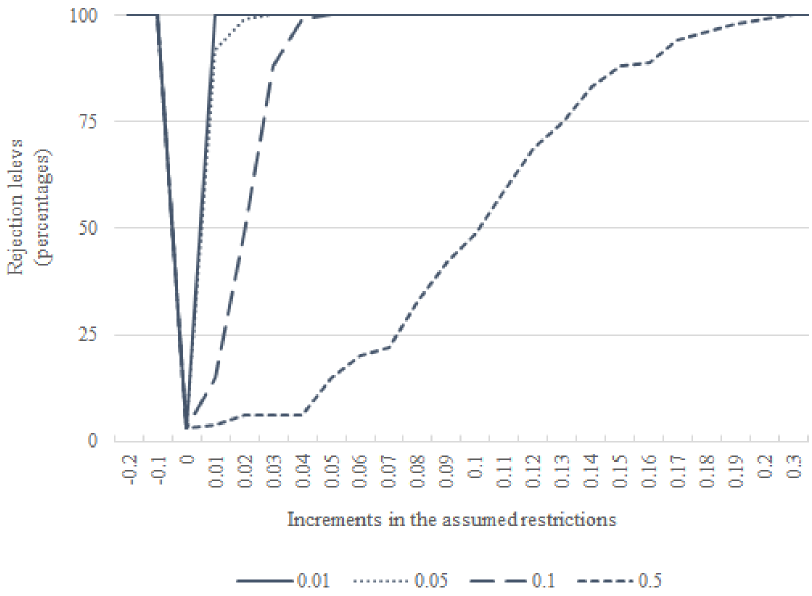

FIGURE 4.2: Wald test power curves for different values of σ_ϵ

4.3.1.2 Monte Carlo simulations with three inputs

Let us now consider DGP2 and look at how two substitution parameters influence the size of the test. We test the coefficients obtained by the estimation of (4.3b) with an added disturbance term jointly for homogeneity and separability. The constraints needed to test if L is separable from E and K, which is the maintained separability assumption of the $((E, K), L)$ CES structure, are given by:

$$\begin{cases} a_1 + a_2 + a_3 = 1 \\ a_{11} + a_{12} + a_{13} = 0 \\ a_{12} + a_{22} + a_{23} = 0 \\ a_{13} + a_{23} + a_{33} = 0 \\ a_1 a_{23} + a_2 a_{13} = 0. \end{cases} \tag{4.7}$$

Simulation results for DGP2 are presented in Table 4.14 for both the cases in which the true assumed production function is CES (first column) and CT (second column). We can see that, in line with the two-input case, when DGP2 is built on a CES production function, rejection levels strongly depend on the substitution parameters and the error variance, positively in the first case and negatively in the second. Conversely, when DGP2 is CT, ρ_x and σ_ϵ influence the results while these are invariant to changes in ρ.

		CES						CT					
	$\rho_x \setminus \rho$	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
$\sigma_\epsilon = 0.01$	*-0.9*	43	7	8	46	49	32	6	5	5	6	6	5
	-0.1	7	5	5	9	33	26	6	5	5	6	6	5
	0.1	8	5	5	9	31	25	6	5	5	6	6	5
	0.9	43	8	7	42	50	31	6	5	5	6	6	5
	2	52	27	27	52	56	38	6	5	5	6	6	5
	9	41	28	29	43	48	39	6	5	5	6	6	5
$\sigma_\epsilon = 0.05$	*-0.9*	7	5	5	7	14	24	5	5	5	5	6	5
	-0.1	5	5	5	6	8	18	5	5	5	5	6	5
	0.1	5	5	5	6	9	18	5	5	5	5	6	5
	0.9	8	6	5	7	13	21	5	5	5	5	6	5
	2	15	7	7	13	24	29	5	5	5	5	6	5
	9	22	12	11	20	32	32	5	5	5	5	6	5
$\sigma_\epsilon = 0.1$	*-0.9*	6	5	5	6	7	12	5	5	5	5	6	5
	-0.1	5	5	5	6	6	9	5	5	5	5	6	5
	0.1	5	5	5	6	6	10	5	5	5	5	6	5
	0.9	6	6	5	6	7	12	5	5	5	5	6	5
	2	8	6	5	6	13	16	5	5	5	5	6	5
	9	10	7	7	10	15	22	5	5	5	5	6	5
$\sigma_\epsilon = 0.5$	*-0.9*	4	3	3	3	3	6	5	5	5	5	5	5
	-0.1	5	5	5	5	5	6	5	5	5	5	5	5
	0.1	5	5	5	6	5	6	5	5	5	5	5	5
	0.9	5	5	5	5	6	5	5	5	5	5	5	5
	2	5	5	5	5	6	7	5	5	5	5	5	5
	9	6	6	5	5	6	6	5	5	5	5	5	5

TABLE 4.14: Size of Wald tests on homogeneity and separability (in percentages) with assumed CES functional form (second column) and CT (third column)

In the three-input case, we can conveniently measure the power of the test assuming the null hypothesis to be one of the alternative separability assumptions (corresponding to the nested structures $((E, L), K)$ and $((K, L), E)$, or the unnested (E, K, L)) and look at how many time we correctly reject it. The constraints that are needed to test for the alternative structures are shown in Table 4.15. If results indicate that the test has the appropriate power, then the Wald test can be used with the purpose of identifying the nested structure that provides the best fit. Simulation results are displayed in Table 4.16. The power of the test decreases with the variance of the disturbances as the true deterministic model is only explaining a small part of the output variation. We also observe that the closer the substitution parameters are to each other (i.e. the closer we are to the diagonal), the lower the power of the test. The explanation is given by the fact that the diagonal elements (where $\rho = \rho_x$) correspond to the unnested CES case where all the separability assumptions are simultaneously satisfied.

((E,L),K)	((K,L),E)	(E,K,L)
$a_1 + a_2 + a_3 = 1$	$a_1 + a_2 + a_3 = 1$	$a_1 + a_2 + a_3 = 1$
$a_{11} + a_{12} + a_{13} = 0$	$a_{11} + a_{12} + a_{13} = 0$	$a_{11} + a_{12} + a_{13} = 0$
$a_{12} + a_{22} + a_{23} = 0$	$a_{12} + a_{22} + a_{23} = 0$	$a_{12} + a_{22} + a_{23} = 0$
$a_{13} + a_{23} + a_{33} = 0$	$a_{13} + a_{23} + a_{33} = 0$	$a_{13} + a_{23} + a_{33} = 0$
$a_1 a_{23} + a_3 a_{12} = 0$	$a_3 a_{12} + a_1 a_{23} = 0$	$a_1 a_{23} + a_3 a_{12} = 0$
		$a_1 a_{23} + a_3 a_{12} = 0$

TABLE 4.15: Separability constraints for alternative nested structures

From the simulation results emerges that, when we test for the homogeneity and separability assumptions characterizing the nested CES form $((E, K), L)$, the Wald test is correctly failing to reject the null of a nested CES production function. Moreover, the constraints corresponding to the $((E, L), K)$, $((K, L), E)$ and (E, K, L) structures are rejected in almost all chosen specifications. Therefore, we can conclude that the Wald test is correctly sized and has power.

		((E,L),K)						((K,L),E)						(E,K,L)					
	$\rho_x \backslash \rho$	-0.9	-0.1	0.1	0.9	2	9	-0.9	-0.1	0.1	0.9	2	9	-0.9	-0.1	0.1	0.9	2	9
$\sigma_\epsilon = 0.01$	-0.9	43	100	100	100	100	100	43	100	100	100	100	100	60	100	100	100	100	100
	-0.1	100	5	100	100	100	100	100	5	100	100	100	100	100	5	100	100	100	100
	0.1	100	100	5	100	100	100	100	100	5	100	100	100	100	100	5	100	100	100
	0.9	100	100	100	42	100	100	100	100	100	42	100	100	100	100	100	58	100	100
	2	100	100	100	100	60	100	100	100	100	100	65	100	100	100	100	100	77	100
	9	100	100	100	100	100	40	100	100	100	100	100	40	100	100	100	100	100	54
$\sigma_\epsilon = 0.05$	-0.9	7	100	100	100	100	100	7	100	100	100	100	100	9	100	100	100	100	100
	-0.1	100	5	94	100	100	100	100	5	93	100	100	100	100	5	95	100	100	100
	0.1	100	93	5	100	100	100	100	94	5	100	100	100	100	95	5	100	100	100
	0.9	100	100	100	7	100	100	100	100	100	7	100	100	100	100	100	9	100	100
	2	100	100	100	100	31	100	100	100	100	100	31	100	100	100	100	100	42	100
	9	100	100	100	100	100	34	100	100	100	100	100	33	100	100	100	100	100	47
$\sigma_\epsilon = 0.1$	-0.9	5	100	100	100	100	100	5	100	100	100	100	100	6	100	100	100	100	100
	-0.1	100	5	34	100	100	100	100	5	34	100	100	100	100	5	35	100	100	100
	0.1	100	36	5	100	100	100	100	35	5	100	100	100	100	37	5	100	100	100
	0.9	100	100	100	6	100	100	100	100	100	6	100	100	100	100	100	7	100	100
	2	100	100	100	100	13	100	100	100	100	100	12	100	100	100	100	100	16	100
	9	100	100	100	100	100	23	100	100	100	100	100	23	100	100	100	100	100	31
$\sigma_\epsilon = 0.5$	-0.9	5	18	28	76	97	100	5	17	28	77	97	100	5	19	29	80	98	100
	-0.1	20	5	5	31	78	99	20	5	6	29	78	99	21	6	6	30	81	100
	0.1	31	7	5	19	67	98	31	6	5	18	66	98	32	7	6	19	69	99
	0.9	77	30	20	5	16	73	76	30	19	5	15	74	80	31	21	6	17	79
	2	97	73	61	19	5	23	97	72	61	17	5	20	98	75	63	20	7	24
	9	100	97	95	70	27	6	100	98	96	68	26	6	100	99	97	74	30	7

TABLE 4.16: Wald tests rejection levels (percentages) for different separability assumptions

4.3.1.3 Discriminating between nested structures

Here, we want to investigate further whether Wald tests could be used not only to understand if data are consistent with a CES representation of the production technology, but also to discriminate between nesting alternatives.

We propose the following approach: for each structure, we run separate Wald tests for homogeneity and separability, collect the χ^2 statistics, and check how often the separability assumption $((E, K), L)$ is characterized by the smallest statistic (i.e. how often we fail to reject the $((E, K), L))$ separability more strongly than the others). The results of the Monte Carlo simulations are given in Table 4.17. In all cases where $\rho \neq \rho_x$, the $((E, K), L)$ restriction is the one with the smallest statistic, indicating that this approach is correctly recognising the assumed nested structure. Moreover, Table 4.17 shows that the diagonal entries are all zero and this indicates that the approach is also correctly rejecting the (E, K, L) restrictions, i.e. the nested structure, where $\rho = \rho_x$.

	$\rho_x \setminus \rho$	-0.9	-0.1	0.1	0.9	2	9
$\sigma_\epsilon = 0.01$	-0.9	0	100	100	100	100	100
	-0.1	100	0	100	100	100	100
	0.1	100	100	0	100	100	100
	0.9	100	100	100	0	100	100
	2	100	100	100	100	0	100
	9	100	100	100	100	100	0
$\sigma_\epsilon = 0.05$	-0.9	0	100	100	100	100	100
	-0.1	100	0	99	100	100	100
	0.1	100	99	0	100	100	100
	0.9	100	100	100	0	100	100
	2	100	100	100	100	0	100
	9	100	100	100	100	100	0
$\sigma_\epsilon = 0.1$	-0.9	0	100	100	100	100	100
	-0.1	100	0	83	100	100	100
	0.1	100	81	0	100	100	100
	0.9	100	100	100	0	100	100
	2	100	100	100	100	0	100
	9	100	100	100	100	100	0
$\sigma_\epsilon = 0.5$	-0.9	0	69	78	97	99	100
	-0.1	70	0	42	80	97	100
	0.1	79	42	0	69	94	100
	0.9	97	78	68	0	64	96
	2	100	96	91	67	0	68
	9	100	100	99	94	74	0

TABLE 4.17: Percentages of times the χ^2 statistic from Wald tests is smallest for (E,K),L

From Table 4.17, it emerges that this approach is partly affected by changes in the substitution parameters and the variance of the disturbances: the $((E, K), L)$ specification is more clearly identified when the variance of disturbances is smaller and the difference between ρ and ρ_x is bigger.

Finally, Table 4.18 presents the results obtained using the R^2 approach, as proposed by the existing literature. By comparing this with the results presented above, we can see that the method proposed here performs better.

	$\rho_x \backslash \rho$	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
$\sigma_\epsilon = 0.01$	*-0.9*	28	100	100	100	100	100
	-0.1	100	38	100	100	100	100
	0.1	100	100	34	100	100	100
	0.9	100	100	100	32	100	100
	2	100	100	100	100	29	100
	9	100	100	100	100	100	32
$\sigma_\epsilon = 0.05$	*-0.9*	29	100	100	100	100	100
	-0.1	100	39	99	100	100	100
	0.1	100	98	36	100	100	100
	0.9	100	100	100	34	100	100
	2	100	100	100	100	29	100
	9	100	100	100	100	100	34
$\sigma_\epsilon = 0.1$	*-0.9*	31	100	100	100	100	100
	-0.1	100	40	83	100	100	100
	0.1	100	81	38	100	100	100
	0.9	100	100	100	37	100	100
	2	100	100	100	100	30	100
	9	100	100	100	100	100	37
$\sigma_\epsilon = 0.5$	*-0.9*	34	65	82	98	100	87
	-0.1	79	37	42	80	99	87
	0.1	78	38	38	70	96	82
	0.9	96	79	66	38	68	71
	2	96	96	87	62	31	53
	9	74	70	69	71	54	10

TABLE 4.18: Percentages of times the R^2 statistic from NLS estimations of alternative nested structures is smallest for the $(E, K), L$ one

4.3.2 Maximum likelihood and non-linear tests

Another class of tests that could be used to identify the correct functional form comprises those tests based on the Likelihood principle. These tests are constructed as the difference between two objective functions, calculated respectively under the null and the alternative hypotheses. The corresponding statistic, under the null, is distributed asymptotically as χ^2, with degrees of freedom equal to the number of constraints imposed.

These tests are econometrically more troublesome than Wald tests as they involve the estimation of two separate models, the restricted and unrestricted ones,[13] but have the advantage of being invariant to the formulation of the hypothesis. Nevertheless, their main disadvantage is that they are based on the assumption of normally distributed disturbances, which is very limiting in empirical applications.

Although the Wald test and Likelihood principle tests are asymptotically equivalent when the constraints are linear,[14] the disparity in results can be very big in the non-linear case and when the sample is small.

In the remainder of this section, we exploit two tests based on the Likelihood principle. The first is a non-linear likelihood ratio (LR) test based on the maximum likelihood estimation of the nested models, CT and UT, that represent the restricted and unrestricted model respectively. The second is a test proposed by Davidson and MacKinnon (1993), which can be used with non-linear least squared estimations. The statistic is the following:

$$DM = (1/MSE)(SSR_c - SSR_u) \tag{4.8}$$

where SSR_c and SSR_u are the residual sums of squares of the constrained (CT) and unconstrained (UT) models, and MSE is the mean squared error of the latter. The statistic is distributed as a χ^2 with degrees of freedom equal to the difference between the parameters of the two models.

4.3.2.1 Monte Carlo simulations with two inputs

Results obtained from the two tests produce identical results that we present in Table 4.19.

$\sigma_\epsilon \backslash \rho$	-0.9	-0.4	-0.1	0.1	0.4	0.9	2	9
0.01	9.7	4.4	4.4	4.4	4.4	10.9	27.3	20.9
0.05	4.6	4.3	4.4	4.4	4.3	5.2	9.5	16.4
0.1	4.5	4.4	4.4	4.4	4.4	4.6	6.4	10.5
0.5	4.0	4.0	4.0	4.0	4.0	4.0	5.0	5.0

Table 4.19: Size of the Likelihood Ratio test (percentages) for DGP1

[13] In the context of this chapter, this is particularly true as the constrained model requires a non-linear estimation.

[14] If they use the same estimated error variance.

These are comparable to those of the Wald tests: the rejections levels are approximately constant across all specifications except for high values of the substitution parameter. However, this time the size of the test is generally smaller than the expected 5%. Therefore, we can conclude that in the two-input case the Wald test should be the preferred inference test to investigate whether a CES function is supported by the available dataset.

4.3.2.2 Monte Carlo simulations with three inputs

Since the results from the two tests are equal to each other and very close to those of the Wald test, we relegate them to Appendix Table B.2. Nevertheless, Table 4.20 reports the simulations results on the percentage of times the χ^2 statistic for the $(E, K), L$ restriction derived from the LR tests is smaller than the statistics obtained testing for all the other feasible separability assumptions. The number of times the LR tests are able to discriminate among the nested structures is similar to that of the Wald test, except for the diagonal entries which are all different from zero. Thus, also in the three-input case, the Wald test should be preferred to the Likelihood principle tests.

4.3.3 Estimated linearised Translog

In this section, we look at the bias and preciseness with which CES parameters are estimated using the linearised CES in both DGP1 and DGP2. The first step is to look at the bias that derives from the use of a linearised model, as we did in the Translog case. We expect the MSBs for the CT estimation to be equal or smaller than the Translog one as its parameters are constrained to satisfy part of the CES maintained hypotheses. In this instance, the bias takes on an specific interpretation: as the CT is the linearisation of an arbitrary CES, the bias represents the error resulting from truncating the Taylor series approximation to the second degree. Thus, we can look at the bias as an approximation error and we expect it to increase as we move away from the point in which the Taylor expansion is made (i.e. the further is(are) the substitution(s) parameter from 0).

We define the two-input and three-input CT respectively as

$$\begin{aligned} q_t^{CT} = g(\beta; E_t, K_t) = \ln(\gamma) + \nu\delta \ln(E) + \nu(1-\delta)\ln(K) - 0.5\rho\nu\delta(1-\delta)\ln(E_t)^2 + \\ - 0.5\rho\nu\delta(1-\delta)\ln(K_t)^2 + \rho\nu\delta(1-\delta)\ln(E_t)\ln(K_t) \end{aligned} \tag{4.9}$$

	$\rho_x \backslash \rho$	-0.9	-0.1	0.1	0.9	2	9
$\sigma_\epsilon = 0.01$	-0.9	41	100	100	100	100	100
	-0.1	100	39	100	100	100	100
	0.1	100	100	39	100	100	100
	0.9	100	100	100	39	100	100
	2	100	100	100	100	46	100
	9	100	100	100	100	100	37
$\sigma_\epsilon = 0.05$	-0.9	38	100	100	100	100	100
	-0.1	100	39	100	100	100	100
	0.1	100	100	39	100	100	100
	0.9	100	100	100	26	100	100
	2	100	100	100	100	38	100
	9	100	100	100	100	100	36
$\sigma_\epsilon = 0.1$	-0.9	37	100	100	100	100	100
	-0.1	100	39	81	100	100	100
	0.1	100	82	39	100	100	100
	0.9	100	100	100	38	100	100
	2	100	100	100	100	38	100
	9	100	100	100	100	100	36
$\sigma_\epsilon = 0.5$	-0.9	39	68	78	97	100	100
	-0.1	70	39	42	80	97	100
	0.1	79	42	38	68	95	100
	0.9	97	79	68	39	64	95
	2	100	96	91	67	37	68
	9	100	100	100	94	74	34

TABLE 4.20: Percentages of times the χ^2 statistic from NL test is the smallest for (E,K),L

and

$$\begin{aligned} q_t^{CT} = g(\beta; E_t, K_t, L_t) = {} & \ln(\gamma) + \delta\delta_x\nu\ln(E_t) + \delta\nu(1-\delta_x)\ln(K_t) - (\delta-1)\nu\ln(L_t) + \\ & + 0.5\delta\delta_x\nu(\delta\delta_x\rho - \delta_x\rho + \delta_x\rho_x - \rho_x)\ln^2(E_t) + \\ & + 0.5\delta\nu(\delta_x-1)(\rho(\delta-1)(\delta_x-1) + \delta_x\rho_x)\ln^2(K_t) + \\ & + 0.5(\delta-1)\delta\nu\rho\ln^2(L_t) + \\ & - \delta\delta_x\nu(\delta_x-1)(\delta\rho - \rho + \rho_x)\ln(E_t)\ln(K_t) + \\ & - \delta\delta_x\nu\rho(\delta-1)\ln(E_t)\ln(L_t) + \\ & + \delta\nu\rho(\delta-1)(\delta_x-1)\ln(K_t)\ln(L_t). \end{aligned} \tag{4.10}$$

The MBS is computed as the difference between the fitted value from the CT non-linear least squares estimation (NLS) and the deterministic CES model assumed in the DGP. The MSBs for the different parametrisations are presented in Table 4.21 and 4.22.

$\sigma_\epsilon \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0038
0.5	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0013	0.0048

TABLE 4.21: Mean squared bias from CT estimation in DGP1

$\rho_x \diagdown \rho$	*0.1*	*0.9*	*2*	*9*	*0.1*	*0.9*	*2*	*9*
	$\sigma_\epsilon = 0$				$\sigma_\epsilon = 0.01$			
-0.9	0.0000	0.0001	0.0004	0.0034	0.0000	0.0001	0.0004	0.0033
-0.1	0.0000	0.0000	0.0001	0.0026	0.0000	0.0000	0.0001	0.0026
0.1	0.0000	0.0000	0.0001	0.0026	0.0000	0.0000	0.0001	0.0026
0.9	0.0000	0.0001	0.0004	0.0033	0.0000	0.0001	0.0004	0.0033
2	0.0001	0.0003	0.0010	0.0048	0.0001	0.0003	0.0010	0.0048
9	0.0009	0.0015	0.0030	0.0084	0.0010	0.0015	0.0030	0.0084
	$\sigma_\epsilon = 0.05$				$\sigma_\epsilon = 0.1$			
-0.9	0.0000	0.0001	0.0004	0.0034	0.0001	0.0002	0.0005	0.0034
-0.1	0.0000	0.0000	0.0001	0.0026	0.0001	0.0001	0.0002	0.0026
0.1	0.0000	0.0000	0.0001	0.0026	0.0001	0.0001	0.0002	0.0026
0.9	0.0000	0.0001	0.0004	0.0033	0.0001	0.0001	0.0005	0.0034
2	0.0001	0.0003	0.0010	0.0048	0.0001	0.0004	0.0011	0.0048
9	0.0010	0.0015	0.0030	0.0084	0.0010	0.0016	0.0030	0.0085
	$\sigma_\epsilon = 0.5$							
-0.9	0.0015	0.0016	0.0019	0.0048				
-0.1	0.0015	0.0015	0.0016	0.0041				
0.1	0.0015	0.0015	0.0016	0.0041				
0.9	0.0015	0.0016	0.0019	0.0048				
2	0.0015	0.0018	0.0025	0.0063				
9	0.0024	0.0030	0.0045	0.0099				

TABLE 4.22: Mean squared bias from CT estimation in DGP2

Comparing them with Table 4.3 and 4.4, respectively, we can see that, in both cases, our expectations are met: closer to the approximation point, the MSBs are approximately the same, but the CT performs better at the extremes and when the variance of the disturbances increases (e.g. the CT model fits better the CES model).

Thursby and Lovell (1978) and Hoff (2014) studied how well the linear approximation to a two-input and n-input CES respectively, estimate the corresponding non-linear CES parameters using Monte Carlo simulation with different parametrisations. Since the Translog approximation to the CES is a Taylor series truncated after two terms, it is unable to capture interactions of higher order between inputs and output. Both concluded that, because of this bias, parameters are estimated consistently only when in the neighbourhood of the approximation point and that, whereas the scale and share parameters are generally estimated with a small bias, the estimated efficiency and substitution parameters tend be characterized by a large bias, especially when ρ departs from zero.

For the two-input case we also look at how well the constant elasticity of substitution is estimated using Monte Carlo simulations and DGP2. In Table 4.23 we can observe that for values of ρ equal or greater than 0.9, results are biased with a positive bias for negative values and *viceversa*. Thus, one needs to be careful when resorting to a linearised CES to estimate the constant substitution relationship between inputs.

	10	*1.66667*	*1.1111*	*0.9091*	*0.7143*	*0.5263*	*0.3333*	*0.1*
$\rho_x \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	5.722	1.646	1.111	0.909	0.718	0.548	0.407	0.314
	(0.126)	(0.010)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)
0.05	5.727	1.649	1.113	0.910	0.719	0.548	0.407	0.314
	(0.610)	(0.049)	(0.022)	(0.015)	(0.009)	(0.006)	(0.003)	(0.003)
0.1	5.758	1.654	1.115	0.911	0.719	0.549	0.407	0.314
	(1.226)	(0.099)	(0.045)	(0.030)	(0.019)	(0.011)	(0.006)	(0.005)
0.5	3.840	1.690	1.130	0.918	0.723	0.551	0.408	0.315
	(5.674)	(0.515)	(0.229)	(0.152)	(0.095)	(0.056)	(0.032)	(0.020)

TABLE 4.23: Estimated constant elasticities from CT regression

Our results for a nested CES regarding $\hat{\nu}$, $\hat{\lambda}$ and $\hat{\rho}$ are in line with the findings of previous literature, as shown in Table 4.24. The first estimated parameter is only slightly affected by the increase in the bias due to the linearised model, while the other two are strongly affected both in terms of bias and precision. For what concerns the share parameter, we need to distinguish between the inner and the outer one: $\hat{\delta_x}$ is estimated with very small bias and standard error tend to remain small for any change in the substitution parameters, whereas $\hat{\delta}$ is estimated with a large bias that increases as ρ and ρ_x depart zero and the effect is more accentuated for changes in ρ_x than in ρ. The inner substitution parameter is estimated with a bias that interestingly becomes smaller for large values of ρ when ρ_x is larger than 0.1.

4.4 Second phase: model selection and elasticities distributions

In the previous sections, we analysed different approaches that could be used to test if the CES maintained hypotheses of homogeneity and separability are supported by

$\rho_x \diagdown \rho$	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
	$\hat{\lambda}$						$\hat{\nu}$					
-0.9	1.503	1.502	1.502	1.501	1.490	1.411	1.000	1.000	1.000	1.000	1.000	1.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.006)
-0.1	1.502	1.500	1.500	1.498	1.487	1.408	1.000	1.000	1.000	1.000	1.000	0.999
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
0.1	1.502	1.500	1.500	1.498	1.487	1.408	1.000	1.000	1.000	1.000	1.000	0.999
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
0.9	1.499	1.498	1.498	1.497	1.487	1.408	1.000	1.000	1.000	1.000	1.000	0.999
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.006)
2	1.489	1.490	1.490	1.490	1.480	1.401	1.000	1.000	1.000	1.000	1.000	0.999
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)	(0.008)
9	1.442	1.441	1.442	1.441	1.428	1.344	1.000	1.000	1.000	1.000	1.000	0.999
	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.010)
	$\hat{\rho}$						$\hat{\rho}_x$					
-0.9	-0.841	-0.100	0.100	0.840	1.544	2.484	-0.856	-0.828	-0.821	-0.792	-0.765	-0.724
	(0.007)	(0.005)	(0.005)	(0.007)	(0.012)	(0.033)	(0.010)	(0.007)	(0.008)	(0.011)	(0.018)	(0.046)
-0.1	-0.842	-0.100	0.100	0.841	1.548	2.495	-0.100	-0.100	-0.100	-0.100	-0.099	-0.097
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.029)	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)	(0.038)
0.1	-0.842	-0.100	0.100	0.841	1.548	2.496	0.099	0.099	0.100	0.100	0.101	0.104
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.029)	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)	(0.037)
0.9	-0.841	-0.100	0.100	0.841	1.544	2.483	0.791	0.821	0.828	0.857	0.881	0.915
	(0.007)	(0.005)	(0.005)	(0.007)	(0.012)	(0.033)	(0.010)	(0.008)	(0.007)	(0.010)	(0.016)	(0.041)
2	-0.839	-0.100	0.100	0.839	1.537	2.462	1.369	1.449	1.470	1.542	1.601	1.671
	(0.011)	(0.006)	(0.006)	(0.010)	(0.017)	(0.039)	(0.017)	(0.010)	(0.010)	(0.015)	(0.024)	(0.049)
9	-0.836	-0.100	0.099	0.837	1.524	2.426	2.029	2.171	2.207	2.324	2.409	2.488
	(0.021)	(0.016)	(0.016)	(0.020)	(0.029)	(0.055)	(0.037)	(0.027)	(0.026)	(0.031)	(0.040)	(0.063)
	$\hat{\delta}$						$\hat{\delta}_x$					
-0.9	0.511	0.501	0.499	0.489	0.480	0.467	0.500	0.500	0.500	0.500	0.500	0.500
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
-0.1	0.501	0.500	0.500	0.499	0.498	0.496	0.500	0.500	0.500	0.500	0.500	0.500
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
0.1	0.499	0.500	0.500	0.501	0.502	0.504	0.500	0.500	0.500	0.500	0.500	0.500
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
0.9	0.489	0.499	0.501	0.511	0.520	0.533	0.500	0.500	0.500	0.500	0.500	0.500
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)
2	0.478	0.497	0.503	0.522	0.540	0.563	0.500	0.500	0.500	0.500	0.500	0.500
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.005)
9	0.455	0.495	0.505	0.545	0.580	0.624	0.500	0.500	0.500	0.500	0.500	0.500
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)

TABLE 4.24: Estimated CES parameters from CT regression

a given dataset. However, inferential tests do not allow us to distinguish between a linearised and a non-linear CES. In order to investigate this matter, we propose two complementary approaches: a graphical analysis on Translog point elasticities distributions, and model selection criteria. The first investigates graphically whether constant elasticities are supported by the dataset; the second provides a formal way of detecting which rival model provides the best representation of the unknown production function.

The graphical approach is based on the distribution of the CT point elasticities of substitution.[15] Indeed, while the CES is characterized by constant elasticities, the CT allows a different degree of substitutability at each inputs and output level. Therefore, we can exploit the Translog estimated coefficients to derive a distribution for each elasticity of substitution. Since the CT elasticities depend on the quantities of inputs, we cannot compute their confidence intervals. However, we can build a prediction interval around each point elasticity which indicates in which range an estimated elasticity of substitution obtained from a new level of inputs and output quantities should fall 95% of the times. For example, suppose that a researcher has data on different industrial sectors for the same year: the prediction interval will inform him on the range in which a new observation for a particular sector will fall 95% of the times.

The graphical analysis of the distribution provides interesting insights: if we observe that the point elasticities are all concentrated around a limited range of values (i.e. we observe a clear peak in the distribution) this is *per se* an evidence that the dataset supports a constant elasticity. Furthermore, the analysis of prediction intervals provides further intuition in support or against the idea of an underlying constant true elasticity: if the intervals are narrow (and similar to one another), it means the point elasticities for the different levels of inputs and output are well predicted and not expected to vary much; on the contrary, if the intervals are wide, it could be an indication that the true production function is not a CES. The best way to appreciate the information that the elasticities distributions and the prediction intervals provide is by means of a graph.

Among the alternative available definitions of elasticity of substitution, we consider the Allen Elasticities of Subsitution (AES) and the Hicks Elasticities of substitution (HES). In a two-input case, the choice of which elasticity should be used is irrelevant, as the different types coincide. Conversely, in a three-input framework we expect HES for the inner nest to be constant, but the other elasticities are allowed to vary. Moreover, in the three-input

[15]Before presenting the graphical analysis, we show the median values of the point elasticities of substitution, and the relative standard error, obtained from Monte Carlo simulations using different parametrisations. This will provide insights on whether the median estimated elasticities are close to the ones defined in the DGPs, and on the magnitude of the standard errors of the predicted median elasticity (to understand how precise is the prediction).

case, the elasticities and their distributions can be used to discriminate between nested structures. Indeed, Christensen and Berndt (1973) show that separability assumptions can be written in terms of AES: if the AES between input E and L and between K and L are equivalent, the $((E, K), L)$ separability assumption is satisfied and, thus, the corresponding nested structure can be deemed appropriate. Given the elasticities between the three inputs, we use a graphical analysis and compare numerical estimates to assess if at least two of them are not statistically and significantly different.

AES are defined as:

$$\sigma_{ij}^{AES} = \frac{\sum_{k=1}^{n} f_k x_k}{x_i x_j} \frac{|D_{ij}|}{|D|}$$

where x_i and x_j are two inputs (e.g. $x_i = E$ and $x_j = K$), f_i, f_j, f_{ii}, and f_{ij} are the first and second partial derivatives of the production function with respect to input x_i and x_j respectively, $|D|$ is the determinant of the bordered Hessian matrix D formed by the estimated coefficients and $|D_{ij}|$ represents the cofactor of the ikth term in the Hessian matrix. HES are defined by:

$$\sigma_{ij}^{HES} = \frac{(f_i x_i + f_j x_j)}{x_i x_j} \frac{f_i f_j}{(2 f_{ij} f_i f_j - f_{jj} f_i^j - f_{ii} f_j^2)}.$$

The second approach is based on model selection criteria. These can be used to choose the "best" model between the CT and the CES. The most employed criteria with non-nested non-linear models are MSE, Aikake information criteria (AIC), and Bayesian information criteria (BIC). Hence, in the following, we run Monte Carlo simulations to look at how frequently these criteria are smaller in the CES versus the CT estimation, which would indicate that the selection criteria are correctly identifying the CES production function as the "true" one.

4.4.1 Graphical analysis

Table 4.25 reports the median values for the Translog estimated point elasticities of substitution with the relative standard error of the prediction.[16] Although in general the median values are overestimated, they are close to the assumed elasticity when ρ is smaller than 0.9 in absolute terms and when the variance of the error term is smaller or equal than 0.1. Once we move away from these specifications, the median estimated elasticity becomes increasingly biased and surrounded by large prediction intervals. In particular, we

[16] The values are obtained from 100 repetitions as each simulation is computationally intensive.

can see that when substitutability is high, the estimated degree of substitutability is more biased than the lower ones.

σ	*10*	*1.6667*	*1.1111*	*0.9091*	*0.7143*	*0.5263*	*0.3333*	*0.1*
$\rho_x \diagdown \rho$	*-0.9*	*-0.4*	*-0.1*	*0.1*	*0.4*	*0.9*	*2*	*9*
0.01	6.240	1.656	1.114	0.912	0.717	0.540	0.379	0.261
	(0.195)	(0.010)	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.004)
0.05	6.305	1.671	1.119	0.914	0.720	0.542	0.380	0.262
	(0.954)	(0.052)	(0.023)	(0.015)	(0.010)	(0.006)	(0.004)	(0.005)
0.1	6.603	1.691	1.127	0.920	0.724	0.544	0.382	0.264
	(2.157)	(0.106)	(0.046)	(0.030)	(0.019)	(0.012)	(0.008)	(0.008)
0.5	2.527	1.973	1.226	0.982	0.761	0.567	0.397	0.275
	(49.20)	(0.768)	(0.276)	(0.174)	(0.106)	(0.064)	(0.042)	(0.034)

TABLE 4.25: Median elasticities of substitution from Translog estimation

As we cannot present graphs for each parametrisation, we display only those that we believe are the most informative according to the findings of previous sections.[17] Each of them is obtained generating 1,000 random values for inputs and output and, for this reason, we should not focus on the numerical values of the estimated elasticities but rather look at whether a constant elasticity is plausible.

Figure 4.3 depicts the elasticity distribution and prediction intervals for the case $\rho = 0.1$ and $\sigma_\epsilon = 0.01$, i.e. a parametrisation where the Translog provides a good representation of the CES. The graph in the first quarter shows the upper and lower bounds of the prediction intervals sorted on the point elasticity they wrap which is displayed on the horizontal axis. The graph in the second quarter represents the distribution of the point elasticities in percentages. The graph in the third quarter is a surface plot which combines the two previous graphs. Finally, the fourth graph is equivalent to the third but in the form of a contour plot. From the first two graphs, we can see that the estimated point elasticities range from 0.9111 to 0.9115 and that the prediction intervals are very narrow and perfectly wrap the true parameter (0.91). This is confirmed by the third and fourth graphs where we observe how the point elasticities are concentrated around a very small range of values. Thus, with this parametrisation, the graphical analysis correctly indicates that the elasticity is approximately constant and a CES can be adopted to describe the production function.

Let us now increase the value of the variance of the disturbances. As shown in Figure 4.4, an increase in σ_ϵ leads to a gradual but limited increase in the range of values for the estimated elasticity but, more interestingly, to a clear enlargement of the prediction intervals. This was expected as the standard errors of the prediction directly depend on the

[17]We include one parametrisation in which we are close to the approximation point and others at the extremes.

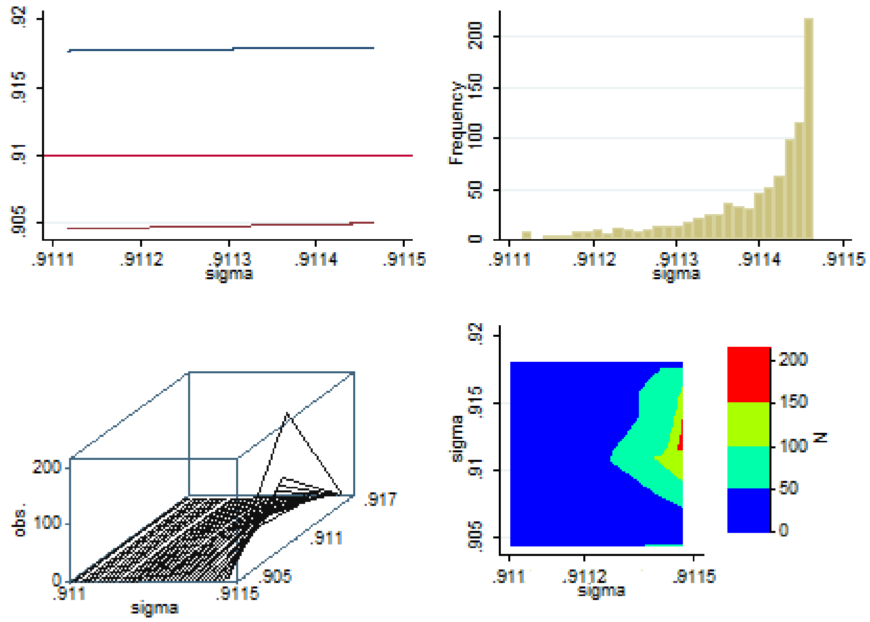

FIGURE 4.3: Point elasticities distribution and prediction intervals with $\rho = 0.1$ and $\sigma_\epsilon = 0.01$ in DGP1

value of the mean squared error which, in turn, depends on the value of the variance of the disturbances. Nevertheless, the range of estimated elasticities is very limited in line with the idea of a CES functional form.

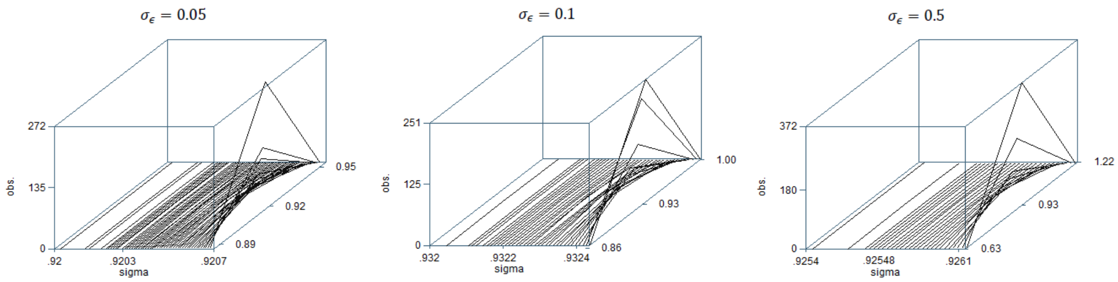

FIGURE 4.4: Surface plots for $\rho = 0.01$ and different values of σ_ϵ in DGP1

Finally, we can look in Figure 4.5 at the effect of a change in the substitution parameter. While the prediction interval remain narrow, the range of point elasticities increases with ρ: for values of ρ smaller than unity, the range is still limited to less that 0.1 and this is an evidence in favour of a CES, for values larger than unity, the number of values taken on by the estimated elasticities becomes too big to support the idea of a constant elasticity.

In the three-input case, results are similar with the only difference that the bias is generally larger because of the sum of the effect from a contemporaneous change in ρ and ρ_x. For

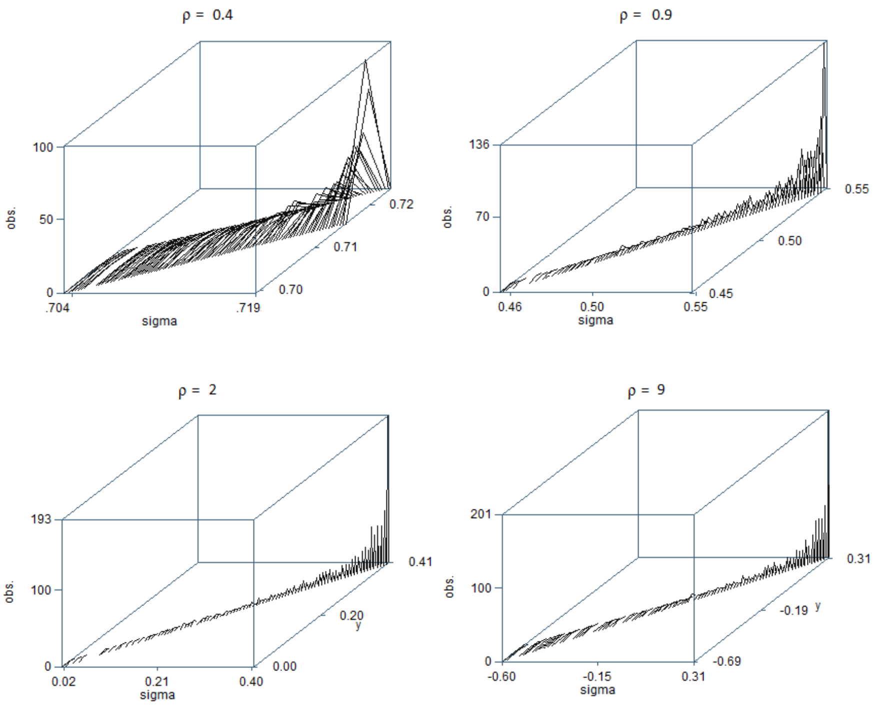

FIGURE 4.5: Surface plots for $\sigma_\epsilon = 0.01$ and different values of ρ in DGP1

this reason, we only look at those cases in which the substitution parameters are close and far. In particular, we focus on the graphs relative to the Allen substitution elasticities between the input outside the nest and each of the two inputs inside as they are informative on the nested structure: if the inference tests results were correct, these should look alike.

Indeed, as Figure 4.6 and 4.7 illustrate, the AES between E and L and K and L look alike and range across the same values while the E-K elasticity is clearly distinguishable. This represents an additional evidence in favour of the $((E, K), L)$ nested structure where the Allen substitution elasticities between the inner inputs and the outer are identical .

4.4.2 Model selection criteria

The results from Monte Carlo simulations for DGP1 for the three model selection criteria are presented in Table 4.26.

They show that the MSE is performing better than the other criteria. Furthermore, the number of times the CES model has the smaller value for the criteria decreases with the increase in the variance of the disturbances. From these formal results, we can correctly conclude that in DGP1 the model that best represents the production function is a CES.

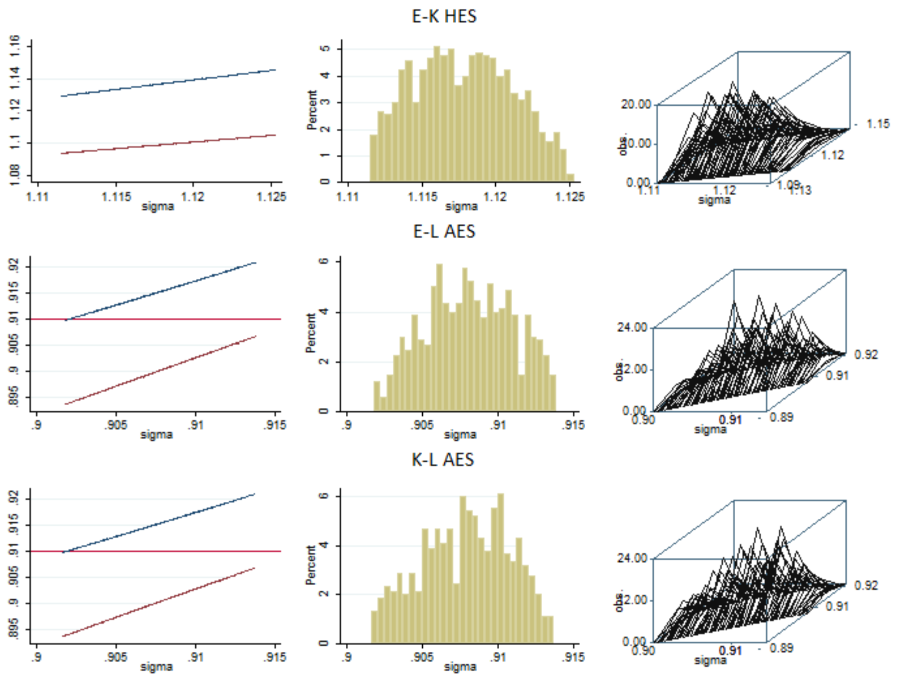

FIGURE 4.6: Point elasticities distributions for $\sigma_\epsilon = 0.01$, $\rho = 0.1$ and $\rho_x = -0.1$. E-K are HES, E-L and K-L are AES

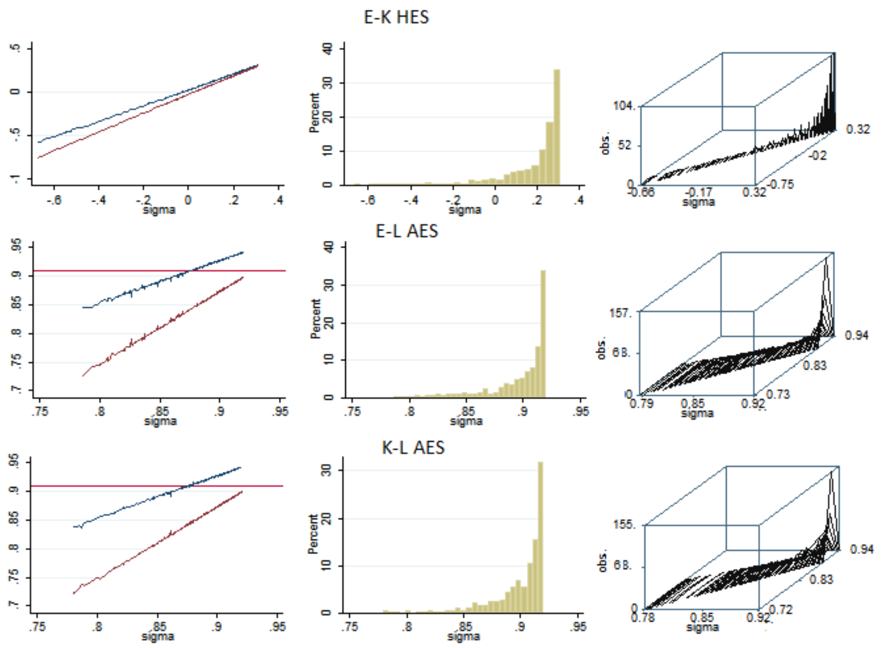

FIGURE 4.7: Point elasticities distributions for $\sigma_\epsilon = 0.01$, $\rho = 0.1$ and $\rho_x = 9$. E-K are HES, E-L and K-L are AES

Results for the three-input case are comparable and we present them in Table B.3 of Appendix B.

	$\rho_x \diagdown \rho$	-0.9	-0.4	-0.1	0.1	0.4	0.9	2	9
MSE	0.01	100	100	100	100	100	100	100	100
	0.05	93	97	100	100	97	93	100	100
	0.1	77	90	100	100	92	77	100	100
	0.5	56	68	89	89	70	54	69	99
AIC	0.01	100	71	85	85	71	100	100	100
	0.05	84	59	97	96	58	84	100	100
	0.1	70	62	96	98	62	70	100	100
	0.5	59	78	96	95	80	58	70	99
BIC	0.01	100	71	84	87	71	100	100	100
	0.05	83	59	96	97	57	84	100	100
	0.1	70	61	97	97	60	70	100	100
	0.5	58	78	95	95	81	58	70	99

TABLE 4.26: Percentages of times selection criteria are smallest for the CES model

4.4.3 Estimated CES function

If the outcome of second phase recommends the use of a CES production function, the best approach for a researcher is to estimate it directly using a NLS regression. Indeed, we observed how the estimated coefficients obtained from the CT regressions become biased when moving away from the approximation point. In Table 4.27 we present the estimated CES constant elasticities obtained from a NLS regression. The CES parameters estimates for DGP1 are reported in Table 4.28. We can see that estimated substitution elasticity bias is very small for low values of the assumed elasticity while the bias quickly increase with the variance of the error term for negative values of σ. The other parameters are estimated with a very small bias across all parametrisations with standard errors that increase with σ_ϵ.

σ	10	1.667	1.111	0.909	0.714	0.526	0.333	0.100
$\rho_x \diagdown \rho$	-0.9	-0.4	-0.1	0.1	0.4	0.9	2	9
0.01	10.030	1.667	1.111	0.909	0.714	0.526	0.333	0.100
	(0.475)	(0.011)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
0.1	10.071	1.675	1.115	0.911	0.715	0.527	0.334	0.100
	(5.050)	(0.108)	(0.045)	(0.030)	(0.020)	(0.013)	(0.010)	(0.012)
0.3	4.955	1.694	1.123	0.914	0.717	0.528	0.334	0.099
	(9.817)	(0.329)	(0.137)	(0.091)	(0.059)	(0.039)	(0.029)	(0.035)
0.5	2.934	1.706	1.131	0.918	0.719	0.530	0.335	0.102
	(7.734)	(0.560)	(0.230)	(0.153)	(0.099)	(0.066)	(0.049)	(0.058)

TABLE 4.27: Estimated constant elasticity from NLS regression of CES as in DGP1

$\rho_x \diagdown \rho$	*0.1*	*0.4*	*0.9*	*2*	*9*	*0.1*	*0.4*	*0.9*	*2*	*9*
			$\hat{\lambda}$					$\hat{\nu}$		
0.01	1.500	1.500	1.500	1.500	1.500	1.000	1.000	1.000	1.000	1.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
0.05	1.500	1.500	1.500	1.500	1.500	1.000	1.000	1.000	1.000	1.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
0.1	1.500	1.500	1.500	1.500	1.500	1.000	1.000	1.000	1.000	1.000
	(0.006)	(0.006)	(0.006)	(0.007)	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
0.5	1.500	1.500	1.500	1.502	1.501	1.000	1.000	0.999	0.999	1.002
	(0.029)	(0.029)	(0.030)	(0.034)	(0.051)	(0.045)	(0.045)	(0.045)	(0.045)	(0.042)
			$\hat{\delta}$					$\hat{\rho}$		
0.01	0.500	0.500	0.500	0.500	0.500	0.100	0.400	0.900	2.000	9.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.004)	(0.004)	(0.005)	(0.009)	(0.117)
0.05	0.500	0.500	0.500	0.500	0.500	0.100	0.400	0.900	1.998	8.999
	(0.002)	(0.002)	(0.002)	(0.004)	(0.008)	(0.014)	(0.014)	(0.015)	(0.020)	(0.456)
0.1	0.500	0.500	0.500	0.500	0.500	0.098	0.398	0.898	1.996	8.998
	(0.004)	(0.005)	(0.005)	(0.006)	(0.016)	(0.036)	(0.038)	(0.047)	(0.088)	(1.174)
0.5	0.501	0.500	0.500	0.501	0.505	0.089	0.390	0.888	1.983	8.805
	(0.023)	(0.023)	(0.025)	(0.030)	(0.082)	(0.183)	(0.194)	(0.237)	(0.437)	(5.587)

TABLE 4.28: CES estimated parameters from a NLS regression with DGP1

Monte Carlo results for the estimated outer and inner elasticities of the nested CES are presented in Table 4.29 and Table 4.30. They are in line with what described in the two-input case: both elasticities tend to be less biased for positive values. Results regarding the nested CES parameters are reported in Table B.4 in the Appendix.

4.5 Conclusions

In this chapter, we proposed a new empirical procedure that can be used to understand if the unknown production function for a given dataset is consistent with a CES and that, when there are more than two inputs, can also be exploited to discriminate against alternative nested structures. This could be of particular interest for researchers attempting the estimation of constant elasticities of substitution to inform CGE model.

We looked at various inference tests that can be used to understand if data support an homogeneous and approximate separable functional form, i.e. a linearised CES. Moreover, we showed how tests on alternative separability assumptions can inform on which nested is closer to the population one. We conclude that the test that performs better in terms of size and power is the Wald test.

σ	*10*	*1.111*	*0.909*	*0.526*	*0.333*	*0.100*	*10*	*1.111*	*0.909*	*0.526*	*0.333*	*0.100*
$\rho_x \diagdown \rho$	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
	$\sigma_\epsilon = 0.01$						$\sigma_\epsilon = 0.05$					
-0.9	10.066	1.112	0.909	0.526	0.333	0.100	10.334	1.114	0.911	0.527	0.334	0.100
	(0.595)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(3.169)	(0.029)	(0.020)	(0.008)	(0.006)	(0.006)
-0.1	10.057	1.112	0.909	0.526	0.333	0.100	10.292	1.113	0.910	0.527	0.334	0.100
	(0.600)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(3.138)	(0.030)	(0.020)	(0.008)	(0.006)	(0.006)
0.1	10.048	1.112	0.909	0.526	0.333	0.100	10.242	1.113	0.910	0.527	0.334	0.100
	(0.600)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(3.121)	(0.030)	(0.020)	(0.008)	(0.006)	(0.006)
0.9	10.017	1.111	0.909	0.526	0.333	0.100	10.066	1.112	0.910	0.527	0.334	0.100
	(0.590)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(3.013)	(0.029)	(0.020)	(0.008)	(0.006)	(0.006)
2	10.037	1.111	0.909	0.526	0.333	0.100	10.182	1.112	0.909	0.526	0.334	0.100
	(0.574)	(0.006)	(0.004)	(0.002)	(0.001)	(0.001)	(2.954)	(0.029)	(0.019)	(0.008)	(0.006)	(0.006)
9	10.013	1.111	0.909	0.526	0.333	0.100	10.048	1.112	0.910	0.526	0.333	0.100
	(0.548)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)	(2.746)	(0.027)	(0.018)	(0.008)	(0.005)	(0.006)
	$\sigma_\epsilon = 0.1$						$\sigma_\epsilon = 0.5$					
-0.9	9.745	1.116	0.913	0.528	0.334	0.100	2.280	1.138	0.927	0.532	0.337	0.104
	(6.851)	(0.060)	(0.040)	(0.016)	(0.011)	(0.012)	(6.077)	(0.308)	(0.204)	(0.083)	(0.056)	(0.058)
-0.1	9.711	1.116	0.912	0.527	0.334	0.099	2.193	1.131	0.922	0.531	0.336	0.106
	(6.732)	(0.060)	(0.040)	(0.017)	(0.011)	(0.012)	(5.984)	(0.312)	(0.206)	(0.083)	(0.057)	(0.058)
0.1	9.619	1.115	0.911	0.527	0.334	0.099	2.191	1.130	0.918	0.529	0.335	0.104
	(6.586)	(0.060)	(0.040)	(0.016)	(0.011)	(0.012)	(5.990)	(0.311)	(0.206)	(0.083)	(0.057)	(0.058)
0.9	9.695	1.114	0.910	0.527	0.334	0.100	2.261	1.125	0.916	0.529	0.334	0.107
	(6.044)	(0.059)	(0.040)	(0.016)	(0.011)	(0.012)	(5.971)	(0.302)	(0.200)	(0.082)	(0.056)	(0.057)
2	9.627	1.113	0.910	0.526	0.334	0.100	2.364	1.117	0.914	0.527	0.335	0.109
	(6.095)	(0.057)	(0.038)	(0.016)	(0.011)	(0.012)	(6.093)	(0.288)	(0.193)	(0.080)	(0.056)	(0.057)
9	9.710	1.112	0.910	0.526	0.333	0.100	2.491	1.115	0.912	0.522	0.333	0.108
	(5.477)	(0.054)	(0.036)	(0.015)	(0.011)	(0.012)	(6.095)	(0.272)	(0.181)	(0.076)	(0.054)	(0.058)

TABLE 4.29: Estimated outer elasticity of substitution from NLS estimation with DGP2

Moreover, once one fails to reject a linearised CES, we illustrated that both a graphical and a formal method can be used to investigate whether the underlying input-output relationship is better represented by a non-linear CES. The graphical method is based on the observation of the distributions of the non-constant estimated substitution elasticities that characterized the linearised CES and the prediction intervals for them: if they range across a limited number of values, we find evidence in favour of a constant elasticity. Conversely, formal tests consist in a comparison between various selection criteria.

Using a Monte Carlo simulation framework where the production function model is assumed to be a CES (or nested CES), we found that the proposed procedure leads to the conclusion that the CES is indeed the most indicated functional form to describe the data in almost all parametrisations.[18]

[18] An exception is represented by the case where the assumed variance of the disturbances is very large.

σ_x	*10*	*1.111*	*0.909*	*0.526*	*0.333*	*0.100*	*10*	*1.111*	*0.909*	*0.526*	*0.333*	*0.100*
$\rho_x \diagdown \rho$	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*	*-0.9*	*-0.1*	*0.1*	*0.9*	*2*	*9*
	$\sigma_\epsilon = 0.01$						$\sigma_\epsilon = 0.05$					
-0.9	10.017	10.000	9.999	9.996	9.977	10.018	9.973	9.806	9.740	9.663	9.577	9.972
	(0.847)	(0.936)	(0.956)	(0.989)	(0.966)	(0.842)	(4.333)	(4.707)	(4.798)	(4.985)	(4.767)	(4.288)
-0.1	1.111	1.111	1.111	1.111	1.111	1.111	1.112	1.112	1.113	1.112	1.111	1.109
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.043)	(0.044)	(0.045)	(0.044)	(0.041)	(0.035)
0.1	0.909	0.909	0.909	0.909	0.909	0.909	0.910	0.910	0.911	0.909	0.909	0.908
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.004)	(0.029)	(0.030)	(0.030)	(0.029)	(0.027)	(0.022)
0.9	0.526	0.526	0.526	0.526	0.526	0.526	0.527	0.527	0.527	0.526	0.526	0.526
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.014)	(0.013)	(0.013)	(0.012)	(0.010)	(0.009)
2	0.333	0.333	0.333	0.333	0.333	0.333	0.334	0.334	0.334	0.333	0.334	0.333
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.010)	(0.010)	(0.010)	(0.009)	(0.008)	(0.006)
9	0.100	0.100	0.100	0.100	0.100	0.100	0.101	0.100	0.100	0.100	0.100	0.100
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.012)	(0.012)	(0.012)	(0.011)	(0.010)	(0.008)
	$\sigma_\epsilon = 0.1$						$\sigma_\epsilon = 0.5$					
-0.9	7.943	7.673	7.636	7.301	7.311	8.347	1.588	1.491	1.417	1.370	1.444	1.672
	(8.423)	(8.555)	(8.626)	(8.802)	(8.324)	(8.185)	(4.757)	(4.391)	(4.323)	(4.206)	(4.265)	(5.056)
-0.1	1.113	1.113	1.114	1.113	1.110	1.108	1.106	1.097	1.097	1.085	1.087	1.096
	(0.086)	(0.089)	(0.090)	(0.088)	(0.083)	(0.069)	(0.435)	(0.452)	(0.462)	(0.444)	(0.420)	(0.345)
0.1	0.910	0.911	0.912	0.910	0.908	0.908	0.909	0.909	0.908	0.905	0.900	0.904
	(0.059)	(0.060)	(0.060)	(0.057)	(0.053)	(0.045)	(0.299)	(0.303)	(0.306)	(0.292)	(0.267)	(0.224)
0.9	0.528	0.528	0.527	0.526	0.526	0.526	0.532	0.531	0.530	0.524	0.523	0.522
	(0.028)	(0.026)	(0.026)	(0.023)	(0.021)	(0.017)	(0.138)	(0.133)	(0.131)	(0.118)	(0.105)	(0.087)
2	0.335	0.334	0.334	0.334	0.334	0.333	0.340	0.336	0.336	0.336	0.334	0.334
	(0.021)	(0.020)	(0.019)	(0.017)	(0.015)	(0.013)	(0.108)	(0.103)	(0.100)	(0.090)	(0.078)	(0.064)
9	0.101	0.100	0.100	0.100	0.100	0.100	0.126	0.123	0.124	0.121	0.119	0.113
	(0.024)	(0.024)	(0.023)	(0.022)	(0.020)	(0.016)	(0.113)	(0.111)	(0.109)	(0.100)	(0.090)	(0.076)

TABLE 4.30: Estimated outer elasticity of substitution from NLS estimation with DGP2

With reference to the procedure presented in this chapter, future research could focus on additional parametrisations. Indeed, although we verified that results are invariant to changes in the efficiency and scale parameters, we observed that changes in the share parameter(s) and in the variance of the input have an impact on the mean squared bias of the model.

More generally, future works should focus on non-nested non-linear hypothesis tests that could be used *a priori* to test between a CES and a Translog production function. Although these tests are discussed in a theoretical framework (Davidson and MacKinnon, 1981, Voung, 1989), we are not aware of any empirical application which exploited them. Another approach may consist in Bayesian analysis where the Translog and the CES functional forms are compared according to diffuse priors.

Chapter 5

Are Elasticities of Substitution Constant?

Empirical evidence using UK production data

5.1 Introduction

The elasticity of substitution of pruduction is defined as the ease with which pairs of inputs can be substituted for one another. The economic literature has been debating over the value and nature of the substitution relationship between energy and capital for a long time and nowadays the topic is still relevant for policy interventions, for instance on energy consumption and emissions reduction. Indeed, substitution elasticity provides information on how costly it is to reduce energy consumption through the introduction of new capital (e.g. new energy-saving machineries), or any other investment able to improve the production process. In fact, when the level of substitution between energy and capital is low, the quantity of capital needed to reduce energy consumption is high if holding output constant. Hence, firms would consider buying new less energy-requiring technology and invest in innovation.[1] A similar argument can be made for what concerns emissions: a low elasticity between energy and capital implies that the costs for being compliant with the established emission targets will be higher in terms of output.

One of the major criticisms of literature on Computable General Equilibrium (CGE) is that the substitution parameters used in the models often lack an empirical foundation and are assumed *a priori* or borrowed from previous studies. However, the value of these parameters can significantly affect the results of the simulations and, as a consequence, the economic insights that can be derived from them. In particular, it has been shown how the substitution elasticities between inputs of production play a crucial role in the energy/environmental CGE models. Saunders (2000), Allan et al. (2007), and Turner (2009) demonstrates how energy use and the size of rebound effects in production are strongly sensible to variations in their value. To address this concern, in this chapter, we

[1] Another possibility for multi-sector or multi-product firms would be to reduce the production of the energy intensive sector/products to the low-intensity ones.

focus on the estimation of the elasticities of substitution using data on multiple industrial sectors for the United Kingdom and a production function consisting of four inputs (i.e. capital, labour, energy, and materials).

Although flexible functional forms (FFF) are sometimes used in CGE models to describe production functions (Despotakis and Fisher, 1988, Li and Rose, 1995, Hertel and Mount, 1985), the great majority of the studies which include at least three factor inputs exploit nested CES functions (see Perroni and Rutherford, 1995). The choice is due to the convenient characteristics and greater tractability of these functional forms: they satisfy the regularity conditions by construction guaranteeing the convergence of the numerical solution of CGE optimization procedures, they are easy to model because their substitution elasticities do not vary with input and output quantities, and yet they allow a certain degree of flexibility as it is possible to specify different pairwise substitution elasticities at each nest.

The empirical literature on substitution elasticities estimation is extensive, from the early work of Berndt and Wood (1975) to the more recent Zha and Ding (2014) and Haller and Hyland (2014), and it is usually based on a FFF cost function, i.e. the Translog, due to the ease with which its share equations and Allen elasticities can be derived. However, as Translog functions are characterized by elasticities that vary with inputs and output quantities, neither the results nor the estimation method can be exploited in a CGE framework.

Unfortunately, the number of papers that estimate nested CES functions to obtain the value of the elasticities of substitution is still very limited. The earliest are those by Prywes (1986) and Chung (1987), followed by Kemfert (1998) and, later, by van der Werf (2008), Okagawa and Ban (2008), Baccianti (2013) and Koesler and Schymura (2015). All these studies have the common intent of informing a CGE model. However, two main problems have been overlooked so far. Firstly the choice of the functional form should be empirically justified: the CES offers the convenient aforementioned characteristics to the detriment of the fact that it is built on strong maintained hypotheses (i.e. homogeneity and separability) which are seldom satisfied by real datasets. Secondly, the use of a nested CES entails the choice of how to specify nesting relationships between inputs. Lecca et al. (2011) show that the choice of a particular form for the nested CES has a remarkable impact on CGE simulation results. While the first CGE papers empirically estimating elasticities of substitution imposed the nested structure *a priori* (Prywes, 1986, Chang, 1994), Kemfert (1998) tried to discriminate between nesting options using the R^2 statistic and this approach was replicated in all the subsequent studies. Whereas it seems convenient, this method does not have a theoretical foundation. The choice of a particular nested structure should instead reflect the separability relationships between inputs. Moreover, mathematical and econometric literature agree that researchers should refrain from using R^2 statistics to compare non-nested non-linear models.

In this chapter, we apply for the first time the new approach proposed in Chapter 4 because it allows us to cope with the two illustrated issues at the same time. The first phase of this approach is based on a FFF, i.e. Translog, whose estimated coefficients can be exploited to test whether the homogeneity and input (approximate) separability conditions maintained in a nested CES are satisfied by the dataset. This not only sheds light on whether a CES is the appropriate functional form to describe the data we analyse, but also testing for different input separability conditions informs on which nested structure best represents the underlying true functional form. If we cannot reject the CES assumptions, in the second phase we perform a graphical analysis of the non-constant distribution of the Translog elasticities and a formal test to find confirmation of whether a non-linear nested CES is supported by the data. Finally, conditional on the result of the previous phases, we proceed with the non-linear estimation of the recommended nested CES, observe the values of its elasticities of substitution and compare them with those obtained from the Translog estimation.

We base our analysis on the EU-KLEMS database provided by the European Commission. We build a panel dataset composed of 23 industrial sectors followed between 1970 and 2005. As the time component is more developed than the number of cross-sectional observations, we correct for multiple econometric issues that are common to this panel structure (i.e. stationarity, serial correlation and contemporaneous correlation).

Results from the first phase indicate that a CES might not be appropriate to describe the dataset under analysis. As discussed in the third chapter, this could be due to a large model bias resulting from the estimation of a CES using a log-linear function. We proceed the analysis with the aim of assessing which nested CES would best approximate our dataset and of estimating the relative constant elasticities. We find that the form $((E,K),L,M)$ is the most appropriate to describe the UK production technology with estimated inner outer elasticities of 0.88 and 0.47 respectively.

The structure of the chapter is the following. In Section 5.2, we provide a brief review of the existing literature. Section 5.3, describes the selected data. In Section 5.4, we present the estimation procedure with the relative potential econometric issues. In Section 5.5, we show the results and report the estimated Translog elasticities of substitution. In Section 5.6, we test for the CES functional form and the in following Section 5.7 we estimate it. Finally, Section 5.8 concludes.

5.2 Literature review

The substitution relationship between inputs of production has been largely investigated from the seminal paper of Berndt and Wood (1975). While the initial interest was connected with the sky-rocketing energy prices which followed the oil crisis in the 1973 (e.g. Berndt and Wood (1975), Griffin and Gregory (1976), Pindyck (1979)), the following studies have been justified by issues like the investment in less energy-intensive physical capital and the depletion of fossil fuels and gas reserves (e.g. Ozatalay et al. (1979), Kim and Heo (2013), Haller and Hyland (2014)) or, more recently, by the increasing energy consumption in developing countries (e.g. Zha and Ding (2014), Zha and Zhou (2014)).[2] The common aim has been to assess whether it is possible to substitute energy with other inputs and mitigate the effects of the rise in energy costs on the economic activity. These studies were generally exploiting a Translog functional form for its generality and the fact that it allows a very straightforward derivation of Allen elasticity of substitution.

More recently, CGE researchers contributed to these literature with the aim of empirically informing the elasticities of substitution for the production side of their models. Indeed, the magnitude of the elasticities have been proven to have an impact on simulation results especially for what concerns analyses on energy shocks and rebound effects. The first paper with this purpose was Kemfert's (1998) for Germany whose work was then further developed by van der Werf (2008) who considered twelve European countries and the U.S. and proposed a new method to estimate the nested CES using cost shares. His work was then followed by those of Okagawa and Ban (2008), Koesler and Schymura (2015) and Baccianti (2013). The common trait of these studies is the use of a CES functional form to describe production. Indeed, although flexible functional forms could be used in a CGE framework, the fact that they are not globally regular and that their elasticities vary with inputs and output make them less appealing from a computational standpoint.

Despite the considerable existing literature and the growing interest, findings are mixed even among studies which use the same dataset and functional form, especially for what concerns the energy and capital relationship.[3] Apostolakis (1990), Thompson and Taylor (1995), and Koetse et al. (2008) formulate different hypotheses to justify the discording results. In particular, Apostolakis (1990) proposes as an explanation the use of different data structures, time-series and cross-section, which lead respectively to long or short period elasticity estimates. Thompson and Taylor (1995) try to demonstrate that results converge using the same type of elasticity of substitution (i.e. the Morishima elasticity).

[2] See the first chapter for a comprehensive literature review.

[3] See the famous debate between Berndt and Wood (1975) and Griffin and Gregory (1976).

Koetse et al. (2008), instead, use a meta-analysis conclude that the reasons for diverging results can be found in the different economic context, econometric procedures, and data characteristics. Chapter 2 builds on Koetse et al. (2008) and shows the main differences between using a CES and a Translog production function and Chapter 4 describes a procedure to discriminate between them and to understand which nested structure provides the best representation of the unknown input-output relationship. This helps the reconciliation between the two strands of literature, the pure econometric and the CGE one.

5.3 Description of the data

A common problem to most of previous literature on the estimation of substitution elasticities has been the lack of a reliable source of data. Often, authors were compelled to create their own input prices and volumes indices using national sources and this was giving rise to problems of measurement errors and comparability of results. For many years the majority of applied studies focused on a single country and sector (generally the entire manufacturing sector) with a very small sample size due to the short time-series availability.

Although gradually single countries became more efficient in collecting data on production allowing researchers to develop analyses based on a bigger sample size, the first harmonised database became available only in 2008, when the EU-KLEMS[4] database was released by the European Commission. This was then followed, in 2012, by the World Input-Output Database (WIOD)[5]. The EU-KLEMS provides data on productivity at industrial level for the members of the European Union from 1970 onwards (the length of the time-series differs between states), harmonising data on capital, labour and intermediate inputs from official national sources and input-output tables. The WIOD provides environmental and socio-economic data at industry-level for 27 European countries and 13 other major countries from 1995 to 2009.

As our analysis is based on a production function, we are interested in the quantities of the four inputs and output for the UK. We opt for the EU-KLEMS database as it provides longer time-series and also produces information on volumes of the materials input which is missing in the WIOD database. In particular, we use data from the March 2008 release as they are the most recent ones that include volume indices for the disaggregated intermediate inputs, i.e. energy and materials.

[4] The data series are also publicly available from the EU-KLEMS website (http://www.euklems.net).

[5] The data series are also publicly available from the WIOD website (http://www.wiod.org).

Our dataset is composed of 23 industrial sectors listed according NACE 1 industry classification (see Table C1 in Appendix C) followed for 36 years (1970-2005) for a total of 828 observations. We use Gross Output volume index as our dependent variable as it measures GDP plus intermediate inputs. Capital quantity is represented by the capital services volumes index which is a quality adjusted measure based on the calculation of a capital stock (using the Perpetual Inventory Method) that takes into account the age-efficiency of different asset types. For labour quantity we use labour services volumes index which is also a quality adjusted measure where the number of hours worked are weighted according to skill types. For the quantities of energy and materials, EU-KLEMS provides two volumes indices. Unfortunately, these are not ideal measurements as they are calculated applying shares from the Use tables to the total intermediate input from national account series.[6] All indices base year is 1995.

5.4 Estimation procedure

5.4.1 Analysis of the time-series

Given the finite number of panels and the long time-series component, we begin our econometric analysis checking for stationarity and cointegration of the inputs and output series.[7] Given the panel nature of the data, we use panel unit-root tests to investigate the order of integration of the series. If we find evidence of non-stationarity, the standard regression techniques are biased and we need to find a stationary combination of the series. In recent years, numerous panel unit-root tests have been proposed which are based on the same principles as the well-known Augmented Dickey-Fuller (ADF) or Phillips-Perron (PP) tests but take into account the unobserved heterogeneity component typical of panel data models. In particular, we consider the Fisher type test by Maddala and Wu (1999) that is feasible with a fixed number of panels N and when the time periods T tend to infinity. The Fisher type test performs separate unit-root tests on each panel and then combines the relative p-values to obtain an overall test statistic. The basic autoregressive model on

[6] While Gross Output and real fixed capital stock match across the different databases (EU-KLEMS, WIOD and OECD), data on labour and energy are very different both in values and trends.

[7] In this chapter, we have used Stata 13 by StataCorp (2013) and the following user written programs: Baum et al. (2002), Kleibergen and Schaffer (2007) (see also Hoyos and Sarafidis, 2006a), Hoyos and Sarafidis (2006b), Schaffer (2005), Schaffer and Stillman (2006), Hoechle (2006) (see also Hoechle, 2007b), Baum (2000a), Baum (2000b).

which the test is based can be expressed formally as:

$$y_{it} = \rho_i y_{i,t-1} + z'_{it}\gamma_i + \epsilon_{it} \tag{5.1}$$

where y_{it} is the series under analysis, $i = 1, ..., N$ indexes panels and $t = 1, ..., T$ indexes time. ϵ_{it} is an idiosyncratic stationary error and z_{it} represents panel specific means and a time trend (i.e. the fixed effects). We test the null that $H_0 : \rho_i = 1$ against the alternative $H_a : \rho_i < 1$, e.g. we test that all panels contain a unit-root against the null that at least one panel is stationary.

At this point we have three alternative outcomes: i) the K, L, E, M, Y series[8] are stationary, ii) the K, L, E, M, Q series are trend-stationary, iii) the K, L, E, M, Q series are integrated. In the first case, we can proceed with the formulation of the model, in the second case we can both de-trend the series or include a time trend in the model, in the third case we perform a panel cointegration test such as the one described in Pedroni (2000). If we find evidence of cointegration, we need to use the Fully Modified OLS (FMOLS) estimator, otherwise we need to differentiate the series according to their degree of integration.

5.4.2 Model specification and panel diagnostics

We begin our analysis assuming a Translog structure for the production function. All previous studies based on a Translog opted for the dual cost function as it allows to use a convenient "standard" procedure based on input demand functions to calculate the Allen elasticities of substitution. However, we base our analysis on the production function for two reasons. First, we do not need to impose assumptions on input prices (i.e. homogeneity) and on competitive markets. Second, we consider fewer data series and this reduces the risk of measurement errors. Third, Translog functions are not self-dual.

Our model is described by the following equation:

$$\begin{aligned} ln(Q_{it}) = {} & a_0 + a_1 ln(E_{it}) + a_2 ln(K_{it}) + a_3 ln(L_{it}) + a_4 ln(M_{it}) \\ & + 0.5a_{11} ln^2(E_{it}) + 0.5a_{22} ln^2(K_{it}) \\ & + 0.5a_{33} ln^2(L_{it})^2 + 0.5a_{44} ln^2(M_{it}) \\ & + a_{12} ln(E_{it}) ln(K_{it}) + a_{13} ln(E_{it}) ln(L_{it}) + a_{14} ln(E_{it}) ln(M_{it}) \\ & + a_{23} ln(K_{it}) ln(L_{it}) + a_{24} ln(K_{it}) ln(M_{it}) + a_{34} ln(L_{it}) ln(M_{it}) \\ & + \alpha_i + \epsilon_{it} \end{aligned} \tag{5.2}$$

[8]K is capital, L is labour, E is energy, M is materials and Q is output.

where Q denotes output, α_i are sector fixed effects and ϵ_{it} is the error term. In case of trend-stationary series, we add a time-trend t to equation (5.2).

Our estimation strategy is carried out in three steps. Given the panel structure of our dataset, we first need to assess if an error component structure is appropriate and, in case, which estimator is the most efficient. We initially test whether α_i are jointly different from zero, e.g. we test for a pooled OLS estimator. If we find an indication that industry unobserved heterogeneity should be included in the model, we perform the Hausman-like overidentifying restriction test on the orthogonality conditions proposed by Arellano to choose between a fixed-effect and a random-effect estimator.

In the second step, we test for heteroskedasticity and serial correlation within panels. In the first case, we use a modified Wald test statistic for group-wise heteroskedasticity as proposed by Greene (2008) which is distributed as a χ^2 with N degrees of freedom under the null of no heteroskedasticity. If we reject the null, we impose White-Huber robust standard errors and, because of the panel structure, we also relax the assumption of independently distributed residuals using clustered standard errors. To test for serial correlation, we use a test for panel data proposed by Wooldridge (2002). If we reject the null of no serial correlation, we use Newey-West standard errors since otherwise our t-tests and F-test would be biased.

Finally, as our panel is characterized by a large T and a small N, we test for cross-sectional dependence, i.e. contemporaneous correlation. Indeed, we suspect a certain degree correlation across industrial sectors. We use the Breusch-Pagan Lagrange Multiplier test of independence whose statistic under the null hypothesis is asymptotically distributed as a χ^2 with $N(N-1)/2$ degrees of freedom. If we reject the null, we find that panels are not independent from one another. To confirm this result, we also use the Pasaran Cross-Sectional Dependence test which under the null is distributed as a standardised normal distribution. The presence of contemporaneous correlation between panels leads to efficiency loss for least squares estimation and to invalid statistical inference. Thus, in this case, we can use Driscoll and Kraay (1998) approach that adjusts the standard errors estimates for various forms of cross-sectional and temporal dependence.

5.5 Estimation results

5.5.1 Diagnostic tests results and Translog estimation

As described above, we begin our econometric analysis looking at the five time-series E, K, L, M and Y. In particular, we want to understand whether the series are stationary over time. We run five separate Fisher type unit-root tests based on the augmented Dickey-Fuller test. We consider a number of lags equal to 1, however results are invariant to other lags specifications. Table 5.1 presents four sets of results for each series: the inverse χ^2, the inverse normal transformations, the relative statistics, and p-values with and without a drift. According to Choi (2001), the inverse normal statistic should be preferred because is the one characterized by the best trade-off between size and power. However, when the number of panels is finite, also the inverse χ^2 test can provide a reliable indication on the presence of unit-roots. We can see that the results of both tests when we do not include a drift in the test reject the null hypothesis for all the series apart from the energy one, E. However, when we include a drift (e.g. a linear trend), we reject the null that all panels contain a unit-root in all cases. Hence, we can conclude that the series are trend-stationary and we account for this including a linear time trend in our estimation.

		No Drift		**Drift**	
Series	**Transformation**	**Statistic**	**P-value**	**Statistic**	**P-value**
E	Inv. χ^2	72.2997	0.0079	164.1673	0.0000
	Inv. normal	-1.8503	0.0321	-8.4278	0.0000
K	Inv. χ^2	47.6096	0.4070	96.3217	0.0000
	Inv. normal	4.5852	1.0000	-2.0317	0.0211
L	Inv. χ^2	38.1780	0.7871	107.5631	0.0000
	Inv. normal	2.3134	0.9897	-5.0401	0.0000
M	Inv. χ^2	63.8453	0.0418	142.0422	0.0000
	Inv. normal	0.1367	0.5544	-6.5974	0.0000
y	Inv. χ^2	58.2289	0.1066	135.6058	0.0000
	Inv. normal	0.5050	0.6932	-6.3568	0.0000

TABLE 5.1: Unit-root test results with and without drift

Now, we present the results of the diagnostic tests described in the previous section. Firstly, we test between pooled, random-effect and fixed-effect estimators. We strongly reject the pooled estimator and the results of the Hausman test on the additional orthogonality restrictions imposed by the random effect estimator indicate that we reject the null with a χ^2 statistics of 287.4 and p-value of 0.

Secondly, we test for heteroskedasticity and serial correlation of the idiosyncratic error. In the first case, we find a χ^2 statistic of 969.6 with a p-value of 0, thus we reject the null of homoskedasticity. In the second case, we strongly reject the null of no first order autocorrelation with a F-statistic of 137.4 and a p-value of 0.

Lastly, we test for simultaneous correlation of the error terms first with Breusch-Pagan LM test and then with Pesaran test: in both cases we strongly reject cross-sectional independence (with χ^2 statistics of 1932.1 and 8.28 respectively and with p-values of 0 in both cases).

Given our findings on heteroskedasticity, serial correlation and cross-sectional correlation, we perform an additional Hausman test between pooled and fixed effect which accounts for the fact that a_i and ϵ_{it} are not *iid* but are affected by different forms of temporal and spacial dependence. We follow Hoechle (2007a) and find confirmation that we need to reject a pooled estimator. This is in line with our previous finding, i.e. the fixed effect estimator is the one that should be preferred given the data under analysis.

Table 5.2 reports the coefficients and standard errors from four within regressions. In particular, the first column shows fixed effect results with OLS standard error, the second column with standard error robust to heteroskedasticity, the third column with standard errors robust to heteroskedasticity and serial correlation and the last column with standard errors robust to heteroskedasticity, serial correlation, and cross-sectional correlation.

Given the high correlation between regressors, we suspect a high degree of multicollinearity that is reflected in the high R^2 (0.837) and the not highly significant coefficients.[9] However, the coefficients by themselves are generally meaningless, thus, we are not interested in their single levels of significance. We are more interested in combinations of them. For example, we can look at the marginal product of the four inputs for the average observation of each industrial sector. These are reported in Table 5.3 together with the relative t-statistics. We can see that they, as the theory predicts, are all between 0 and 1 and given the critical value of $t_{.025,35} = 2.03$, most of the marginal products are highly significant with few exceptions for the marginal products of labour (MPL). From Table 5.3 we can see that the marginal product of energy (MPE) and labour do not vary much across the different sectors as opposed to the marginal product of capital (MPK) and materials (MPM). MPL are generally the smallest and MPK the largest. We can also observe that the returns on capital are the largest in the Wood and Cork and in the Electricity sectors and the MPE are bigger in the Mining and Quarrying and Electricity, Gas and Water supply sectors.

[9] To overcome this problem we could have used a Seemingly Unrelated Equations estimation using input cost shares. However, in that case, we cannot correct the variance-covariance matrix for the numerous econometric problems we identified with the diagnostic tests.

Variable	**FE**	**White**	**Newey**	**Driscoll**
$ln(E)$	-0.1900	-0.1900	-0.1900	-0.1900*
	(0.1996)	(0.4657)	(0.2942)	(0.1979)
$ln(K)$	-0.5788	-0.5788	-0.5788	-0.5788
	(0.3239)	(0.9754)	(0.4833)	(0.2456)
$ln(L)$	0.2342	0.2342	0.2342	0.2342
	(0.3212)	(0.8366)	(0.4727)	(0.3629)
$ln(M)$	-0.7846*	-0.7846	-0.7846	-0.7846*
	(0.3350)	(0.6922)	(0.4862)	(0.3816)
$ln(E)^2$	0.0010	0.0010	0.0010	0.0010
	(0.0096)	(0.0410)	(0.0138)	(0.0132)
$ln(K)^2$	0.2028***	0.2028***	0.2028***	0.2028***
	(0.0343)	(0.0968)	(0.0510)	(0.0300)
$ln(L)^2$	-0.0161	-0.0161	-0.0161	-0.0161
	(0.0224)	(0.0778)	(0.0336)	(0.0325)
$ln(M)^2$	-0.1156***	-0.1156	-0.1156**	-0.1156*
	(0.0239)	(0.0762)	(0.0352)	(0.0475)
$ln(E)ln(K)$	-0.2356***	-0.2356	-0.2356***	-0.2356***
	(0.0373)	(0.1280)	(0.0543)	(0.0434)
$ln(E)ln(L)$	0.1053***	0.1053	0.1053*	0.1053**
	(0.0298)	(0.0937)	(0.0436)	(0.0394)
$ln(E)ln(M)$	0.2070***	0.2070	0.2070***	0.2070***
	(0.0262)	(0.1409)	(0.0379)	(0.0587)
$ln(K)ln(L)$	-0.1574***	-0.1574	-0.1574**	-0.1574***
	(0.0361)	(0.0934)	(0.0550)	(0.0424)
$ln(K)ln(M)$	0.2025***	0.2025	0.2025**	0.2025***
	(0.0441)	(0.0848)	(0.0651)	(0.0613)
$ln(L)ln(M)$	0.0580	0.0580	0.0580	0.0580
	(0.0373)	(0.1121)	(0.0553)	(0.0592)
t	-0.0014	-0.0014	-0.0014	-0.0014
	(0.0008)	(0.0026)	(0.0012)	(0.0016)
constant	5.3653***	5.3653*		5.3653***
	(1.1795)	(2.4045)		(1.2525)
R^2	0.836			

* indicates a level of significance of 10%, ** indicates a level of significance of 5%, *** indicates a level of significance of 1%,

TABLE 5.2: Fixed effect estimation with different standard errors (in parenthesis)

Furthermore, we can look at the level of returns to scale of our production function. From the estimated coefficients we obtain a coefficient of returns to scale of 0.542, statistically significant at a 5% level. This indicates that the production function for the UK is characterised by decreasing returns.

As the last step of our estimation results, we have to check whether the Translog is well-behaved, e.g. if output is monotonically increasing and the isoquants are convex. The Translog does not satisfy these conditions globally so we need to test our fitted Translog for monotonicity and convexity at each observation. Monotonicity is guaranteed by positive fitted marginal products. Although many studies on the estimation of elasticities substitution with a Translog function assumed well-behaved production functions without testing for it (Ozatalay et al., 1979, Norsworthy and Malmquist, 1983, Moghimzadeh and Kymn, 1986, Garofalo and Malhotra, 1988, Hisnanick and Kyer, 1995, Christopoulos, 2000, Khiabani and Hasani, 2010, Kim and Heo, 2013), others have verified if their estimated Translog satisfied the regularity conditions. Among these, few found they were satisfied on all the domain (Berndt and Wood, 1975, Griffin and Gregory, 1976, Fuss, 1977, Turnovsky et al., 1982, Burki and Khan, 2004, Roy et al., 2006) but in numerous other cases monotonicity or the curvature conditions were rejected for at least some of the observations in the dataset. The consequent responses have been manifold: exclude all the observations where the monotonicity condition were not satisfied but keep those where isoquants convexity was rejected (Medina and Vega-Cervera, 2001), remove the sectors/countries that were more affected by the rejection (Field and Grebenstein, 1980, Medina and Vega-Cervera, 2001), proceed with the estimation ignoring the rejection (Dargay, 1983, Hesse and Tarkka, 1986, Nguyen and Streitwieser, 1999).

When we test for monotonicity, we find that this property is violated for 107 observations. Then we test for convexity of the isoquants checking whether the Bordered Hessian matrix is negative definite, i.e. the successive principal minors alternate in sign, and find that the condition is not satisfied for the same 107 observations and for other 140. For the remaining of this chapter, we drop the 107 observations violating monotonicity, but we keep the additional 140 that only violate convexity of isoquants, since results are not significantly affected by their inclusions.

5.5.2 Estimated point elasticities

In this section, we calculate the elasticities of substitution between the four factors of production. When the production function is composed by more than two inputs, a number

Sector	MPE	t	MPK	t	MPL	t	MPM	t
Agric., Hunting, Forestry and Fishing	0.168	10.525	0.418	9.327	0.078	1.795	0.270	6.634
Mining and Quarrying	0.359	10.579	0.208	6.755	0.097	1.779	0.183	3.578
Food, Beverages and Tobacco	0.189	10.073	0.416	9.341	0.087	2.139	0.251	7.174
Textiles, Leather and Footwear	0.170	7.524	0.461	9.346	0.082	1.688	0.311	6.396
Wood and Of Wood and Cork	0.151	4.225	0.708	8.805	-0.015	-0.374	0.113	2.539
Pulp, Paper, Printing and Publishing	0.145	8.486	0.326	8.187	0.111	2.058	0.325	5.840
Chemical, Rubber, Plastics and Fuel	0.173	8.381	0.254	6.052	0.138	2.189	0.405	4.917
Other Non-Metallic Mineral	0.205	17.965	0.246	6.917	0.125	2.333	0.342	4.888
Basic Metals and Fabricated Metal	0.170	7.196	0.406	8.408	0.099	1.698	0.381	5.413
Machinery, Nec	0.169	9.719	0.436	9.423	0.075	1.673	0.285	6.438
Electrical and Optical Equipment	0.153	17.073	0.301	8.056	0.083	1.604	0.304	4.989
Transport Equipment	0.212	16.936	0.297	7.839	0.107	2.015	0.339	4.904
Manufacturing Nec, Recycling	0.172	10.148	0.379	8.342	0.071	1.838	0.175	6.174
Electricity, Gas and Water Supply	0.293	6.126	0.526	8.319	0.090	3.045	0.095	2.428
Construction	0.128	6.884	0.441	9.090	0.073	1.442	0.309	6.314
Wholesale and Retail Trade	0.178	11.982	0.312	8.384	0.111	2.483	0.251	6.440
Hotels and Restaurants	0.079	3.289	0.292	6.863	0.102	1.547	0.378	5.072
Transport and Storage	0.112	8.076	0.378	8.526	0.070	1.315	0.314	5.524
Post and Telecommunications	0.167	12.439	0.296	8.025	0.114	2.207	0.320	5.538
Public Adm. and Defence	0.223	9.152	0.371	8.728	0.115	3.091	0.219	8.130
Education	0.180	4.447	0.512	8.463	0.113	2.160	0.240	7.818
Health and Social Work	0.120	4.961	0.231	6.099	0.143	2.168	0.363	5.315
Other Community Services	0.195	5.309	0.348	7.105	0.156	2.776	0.223	7.854

TABLE 5.3: Marginal product for the KLEM inputs with the relative t-statistics

of different definitions of elasticity of substitution have been suggested in the literature. The three most common are the Hicks (or direct) elasticity of substitution (HES), the Allen elasticity of substitution (AES) and the Morishima elasticity of substitution (MES). They differ in economic interpretation and implications. The HES are the direct generalization of the Hicks elasticities to an *n*-input function, when computed between two inputs the remaining input quantities are hold constant. For this reason they are usually seen as short-term elasticities. AES are the most widely estimated elasticities and are characterized by the fact that they span from negative to positive values, indicating complementarity and substitutability respectively. Finally, MES are the most recent definition of elasticity of substitution and Blackorby and Russell (1989) argued that they are the only ones which are able to truly represents the nature of the relationship between inputs. They have the particular feature of being asymmetric.

To simplify comparisons with other studies, we separately compute the three forms of elasticities from the estimated Translog coefficients. Since the Translog production function is characterized by elasticities of substitution that vary with input and output, we are going to find a distribution for each of the six elasticities. In Table 5.4 we report the median HES, AES, and MES.

	HES	**AES**	**MES**
EK	1.106	2.519	1.377
EL	0.556	-4.376	-0.4681
KL	0.293	-0.544	-0.149
EM	1.915	-2.998	-1.325
KM	0.083	-0.039	-0.308
LM	0.188	2.297	0.433

TABLE 5.4: Median values of the HES, AES, MES

We can observe how all three elasticities support energy and capital substitutability. Another interesting result is that we find evidence of capital and labour complementarity. For E-M and L-M we find contradictory results: in the first case, HES indicate that the two inputs are substitutes but in terms of AES and MES they are complements; in the second case HES indicates that the two inputs are complements and AES and MES that they are substitutes.

In Table 5.5, 5.6, and 5.7 we present mean estimated values respectively of the HES, AES, and MES for each industrial sector. We can see that, a part from the K-M elasticities, the sign of the substitution relationships between inputs remains the same across sectors and the magnitude does not vary extensively. If we look at the energy-intensive sectors,[10] we

[10] Agric., Hunting, Forestry and Fishing; Mining and Quarring; Textiles, Leather and Footwear; Wood; Pulp, Paper, Printing and Publishing; Chemical, Rubber, Plastics; Other non-metallic mineral; Basic metals and Fabricated metal; Electricity, Gas and Water Supply; Construction; Transport and Storage.

observe that all of them are characterized by high levels of E-K substitutability: this is good news for environmental policy as, even without technological progress, input substitution have the potential to reduce firms demand for energy without large output losses. The only exception is represented by the Mining and Quarrying sector which shows the lowest E-K elasticity independently from the type of elasticity observed. This indicates that in this particular sector is less easy to substitute the two inputs.

	EK	**EL**	**KL**	**EM**	**KM**	**LM**
Agric., Hunting, Forestry and Fishing	2.913	-3.942	-0.528	-2.637	0.113	1.474
Mining and Quarrying	1.795	-4.340	-0.572	-2.307	0.183	1.039
Food, Beverages and Tobacco	3.174	-4.110	-0.588	-2.887	0.019	1.672
Textiles, Leather and Footwear	2.620	-4.529	-0.535	-2.988	0.054	1.625
Wood and Of Wood and Cork	2.467	-4.762	-0.565	-3.831	-0.180	2.792
Pulp, Paper, Printing and Publishing	2.309	-3.712	-0.595	-2.580	-0.080	2.402
Chemical, Rubber, Plastics and Fuel	2.105	-4.066	-0.473	-2.619	0.145	2.303
Other Non-Metallic Mineral	2.100	-4.502	-0.595	-3.163	-0.046	2.198
Basic Metals and Fabricated Metal	2.227	-4.671	-0.518	-3.248	0.045	2.324
Machinery, Nec	2.685	-4.428	-0.582	-2.631	0.004	1.745
Electrical and Optical Equipment	2.573	-3.702	-0.959	-1.741	0.228	1.419
Transport Equipment	2.118	-3.787	-0.631	-1.856	0.326	1.076
Manufacturing Nec, Recycling	3.424	-4.602	-0.688	-3.397	-0.123	2.217
Electricity, Gas and Water Supply	2.838	-4.401	-0.398	-3.571	-0.147	2.065
Construction	2.584	-4.353	-0.664	-2.928	-0.078	2.391
Wholesale and Retail Trade	2.712	-4.153	-0.641	-2.865	-0.100	1.937
Hotels and Restaurants	1.933	-3.389	-0.439	-2.139	0.096	2.777
Transport and Storage	2.408	-3.683	-0.710	-2.263	-0.050	2.380
Post and Telecommunications	2.558	-3.721	-0.601	-2.077	-0.035	1.905
Public Adm. and Defence	3.123	-3.435	-0.466	-2.481	-0.048	1.334
Education	2.614	-4.607	-0.413	-3.769	-0.541	2.707
Health and Social Work	2.467	-3.253	-0.369	-3.035	-0.073	2.478
Other Community Services	2.411	-3.719	-0.525	-3.637	-0.339	2.771

Table 5.5: Mean estimated Allen elasticities of substitution by sector

5.6 Test for CES

In this section we check whether the data we analyse support a CES production function. As discussed in the fourth chapter, in a first phase we test jointly for homogeneity and approximate separability of inputs using Wald tests. If these conditions are not rejected, in a second phase we use a graphical analysis and model selection criteria to confirm whether a nested CES is appropriate to describe the true underlying input-output relationship.

	EK	EL	KL	EM	KM	LM
Agric., Hunting, Forestry and Fishing	1.175	1.001	0.328	3.152	0.918	0.335
Mining and Quarrying	0.677	0.445	-0.037	0.998	0.819	0.182
Food, Beverages and Tobacco	1.127	1.194	0.367	2.725	0.925	0.360
Textiles, Leather and Footwear	1.129	0.836	0.348	1.758	0.901	0.328
Wood and Of Wood and Cork	1.088	0.835	0.372	1.824	0.771	0.376
Pulp, Paper, Printing and Publishing	1.092	0.251	0.183	1.831	0.889	0.235
Chemical, Rubber, Plastics and Fuel	1.092	0.288	0.024	1.927	0.860	0.345
Other Non-Metallic Mineral	1.095	0.358	0.226	1.996	0.913	0.316
Basic Metals and Fabricated Metal	1.100	0.324	0.187	1.955	0.865	0.256
Machinery, Nec	1.112	0.520	0.230	1.907	0.929	0.302
Electrical and Optical Equipment	1.134	0.217	0.063	2.431	1.014	0.367
Transport Equipment	1.136	0.520	-0.104	2.142	1.162	0.351
Manufacturing Nec, Recycling	1.120	0.962	0.495	2.283	0.791	0.329
Electricity, Gas and Water Supply	1.095	1.481	0.650	1.327	0.776	0.396
Construction	1.092	0.451	0.241	1.876	0.870	0.294
Wholesale and Retail Trade	1.096	0.522	0.282	2.092	0.761	0.306
Hotels and Restaurants	1.141	0.105	0.091	1.612	0.782	0.170
Transport and Storage	1.103	0.089	0.106	1.787	1.003	0.118
Post and Telecommunications	1.108	0.132	0.084	1.899	0.793	0.170
Public Adm. and Defence	1.155	0.828	0.478	2.108	0.934	0.390
Education	1.098	0.822	0.393	1.740	0.730	0.301
Health and Social Work	1.134	0.821	0.388	1.478	0.748	0.358
Other Community Services	1.107	0.911	0.393	1.489	0.722	0.212

TABLE 5.6: Mean estimated Hicks elasticities of substitution by sector

5.6.1 Formal tests

We begin the first phase with a Wald test on homogeneity and show results in Table 5.8. We can see that homogeneity is rejected at 10% level and this is mostly due to the fact that the homogeneity restriction regarding the capital input is strongly rejected. This result is thus indicating that the production function representing the analysed dataset is not consistent with a CES. Nevertheless, we could argue that a CES might be the appropriate model to describe input-output relationship but that the bias resulting from the estimation of a Translog is large and it is affecting test results. Furthermore, the CGE literature would still want to find the constant elasticity/ies that best describes the degree of substitution between inputs for the chosen dataset. In the following, we illustrate further steps that one can take to find those elasticities.

As separability restrictions are different for alternative nested structures, a Wald test on approximate separability allows to discriminate between them. With four inputs, the number of possible nested structures is very large. Especially if we consider nested CES

	EK	EL	KL	EM	KM	LM
Agric., Hunting, Forestry and Fishing	1.447	-0.444	-0.098	-1.191	-0.092	0.456
Mining and Quarrying	1.247	-0.036	-0.232	-1.319	-0.302	0.420
Food, Beverages and Tobacco	1.450	-0.482	-0.137	-1.268	-0.222	0.446
Textiles, Leather and Footwear	1.400	-0.578	-0.139	-1.333	-0.113	0.444
Wood and Of Wood and Cork	1.391	-0.740	-0.261	-1.461	-0.425	0.422
Pulp, Paper, Printing and Publishing	1.324	-0.316	-0.139	-1.364	-0.404	0.465
Chemical, Rubber, Plastics and Fuel	1.288	-0.047	-0.019	-1.334	-0.310	0.542
Other Non-Metallic Mineral	1.269	-0.468	-0.184	-1.415	-0.343	0.468
Basic Metals and Fabricated Metal	1.302	-0.367	-0.125	-1.418	-0.287	0.497
Machinery, Nec	1.407	-0.312	-0.105	-1.411	-0.240	0.436
Electrical and Optical Equipment	1.485	0.252	-0.032	-1.309	-0.115	0.515
Transport Equipment	1.265	0.129	-0.100	-1.189	0.106	0.456
Manufacturing Nec, Recycling	1.497	-0.599	-0.212	-1.370	-0.244	0.422
Electricity, Gas and Water Supply	1.476	-0.832	-0.226	-1.199	-0.247	0.370
Construction	1.376	-0.444	-0.171	-1.427	-0.408	0.496
Wholesale and Retail Trade	1.367	-0.429	-0.166	-1.312	-0.343	0.418
Hotels and Restaurants	1.293	-0.174	-0.071	-1.346	-0.375	0.448
Transport and Storage	1.338	-0.110	-0.043	-1.423	-0.372	0.535
Post and Telecommunications	1.356	-0.124	-0.040	-1.203	-0.318	0.401
Public Adm. and Defence	1.431	-0.578	-0.139	-1.030	-0.098	0.402
Education	1.397	-0.776	-0.200	-1.279	-0.313	0.396
Health and Social Work	1.384	-0.726	-0.212	-1.121	-0.307	0.390
Other Community Services	1.361	-0.681	-0.240	-1.393	-0.375	0.339

Table 5.7: Mean estimated Morishima elasticities of substitution by sector

Null hypothesis	Test	Statistic	p-value
$(a_{11} + a_{12} + a_{13} + a_{14} = 0)$	F(1,35)	2.86	0.10
$(a_{22} + a_{12} + a_{23} + a_{24} = 0)$	F(1,35)	13.18	0.00
$(a_{33} + a_{13} + a_{23} + a_{34} = 0)$	F(1,35)	0.10	0.76
$(a_{44} + a_{14} + a_{24} + a_{34} = 0)$	F(1,35)	6.00	0.02
(All the above)	F(4,35)	16.01	0.00

Table 5.8: Wald tests on homogeneity for different nested structures

functions composed by three levels of production, e.g. $(((K, L), E), M)$. In the following of this section we only present results for the structures that we consider sensible from an economic point of view, i.e. those structures that make economic sense.[11]

Table 5.9 presents the Wald test results for the joint homogeneity and approximate separability assumptions (which we expect to reject) and a test on approximate separability

[11]For example, we do not include the $((E, L), (K, M))$ structure as it would suggest that at a lower level of production energy and labour and capital and materials are combined to form intermediate goods which is highly unrealistic.

alone. Results indicate that among the structures for which we fail to reject the null of separability, the two-level $((E,K),L,M)$ nested CES should be preferred given its smaller χ^2 statistic and considerably larger p-value.

Nested structure	**Test (H&S)**	**Statistic**	**p-value**	**Test (S)**	**Statistic**	**p-value**
(K,L,E,M)	F(8,35)	371.76	0.00	F(4,35)	7.59	0.11
((K,L,M),M)	F(8,35)	841.08	0.00	F(4,35)	9.90	0.04
(K,L),(E,M)	F(7,35)	408.41	0.00	F(3,35)	3.54	0.31
((K,L),E,M)	F(8,35)	198.44	0.00	F(4,35)	8.30	0.08
((E,K),L,M)	F(8,35)	197.67	0.00	F(4,35)	2.71	0.61
(((K,L),E),M)	F(7,35)	204.03	0.00	F(3,35)	6.69	0.08
(((K,L),M),E)	F(7,35)	282.58	0.00	F(3,35)	8.61	0.03
(((E,K),L),M)	F(7,35)	237.40	0.00	F(3,35)	6.48	0.09

TABLE 5.9: Wald tests on homogeneity and separability (H&S) and separability alone (S) for different nested structures

5.6.2 Graphical analysis

Graphical analysis of Translog point elasticities could also provide an indication on how far elasticities are from being constant. This analysis is based on the distribution of the Translog estimated substitution elasticities and on the prediction intervals constructed around each of them. They show the range inside which an estimated elasticities obtained from new values of inputs and output quantities for a certain sector will fall 95% of times.

An important evidence in favour of the CES functional form can be obtained looking at the distribution of the estimated elasticities. If the distribution peaks around few values and is not uniformly distributed, i.e. the elasticity values remain quite stable across the sample, a constant elasticity is supported by the data and, hence, a CES specification. Also, the size of the prediction intervals helps to gauge how much the elasticities vary: if the interval is narrow, a new point elasticity is predicted to fall in that particular precise range.

In the following of this section, we show three graphs for each elasticity: the first graph represents the lower and upper bounds of the interval for each point elasticity, the second shows the elasticities distributions and the third combines the two previous graphs in a surface graph. In this analysis we consider only the HES as they are the ones that are constant in a nested CES function. We control for outliers excluding the highest and lowest 10% of the estimated elasticities.

What emerges from the graphs is that the range of estimated point elasticities is the smallest that is indeed the capital-energy one: estimated elasticities vary from approximately 1.08

and 1.5 but from Figure 5.1 we can see that most of the values lie between 1.08 and 1.3. Moreover, the prediction intervals around those values are quite narrow (the value of the lower and upper bounds of the interval in the interval 1.08 and 1.15 are approximately 0 and 2 respectively) indicating that the point elasticity variation is limited. The surface graph confirms this intuition showing a narrow peak around 1.1. The remaining elasticities show larger variation in the point elasticities distribution. Prediction intervals are in general quite narrow though, indicating that each point elasticities is well predicted. We can conclude that the graphical analysis is in line with the recommendation obtained from the formal nesting tests, i.e. the E-K elasticity is the "most constant".

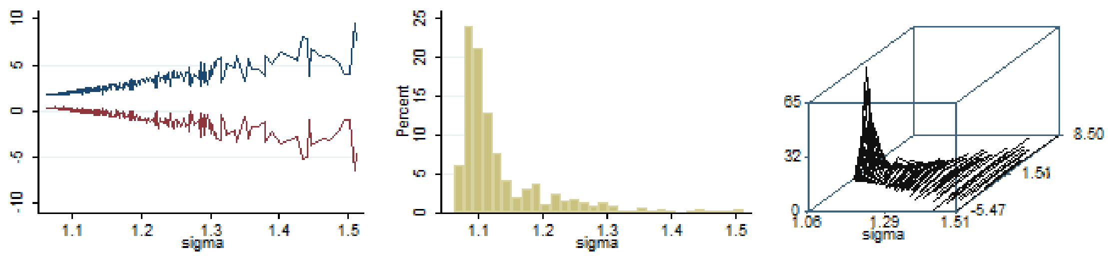

FIGURE 5.1: Translog estimated E-K Hicks elasticities graphical analysis

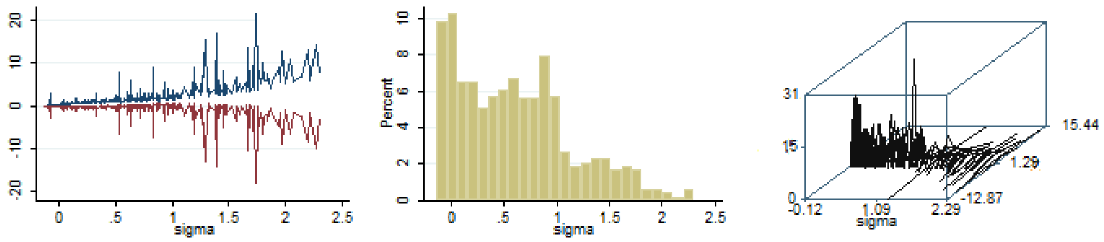

FIGURE 5.2: Translog estimated E-L Hicks elasticities graphical analysis

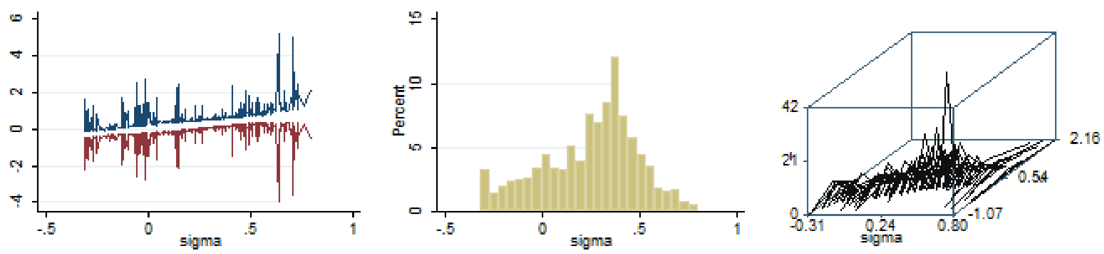

FIGURE 5.3: Translog estimated K-L Hicks elasticities graphical analysis

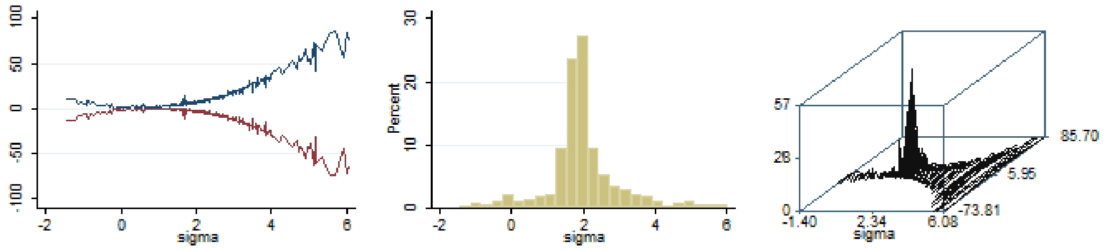

FIGURE 5.4: Translog estimated E-M Hicks elasticity graphical analysis

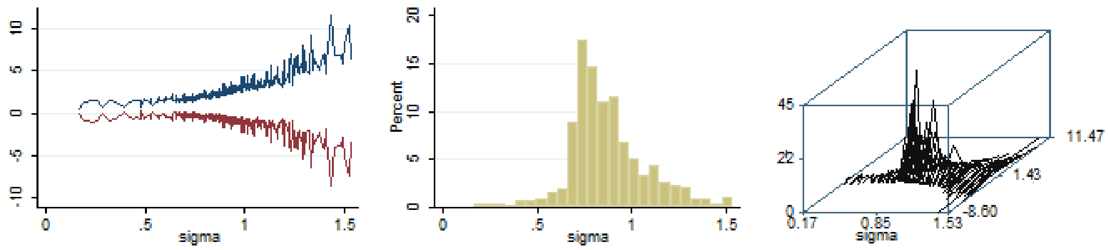

FIGURE 5.5: Translog estimated K-M Hicks elasticity graphical analysis

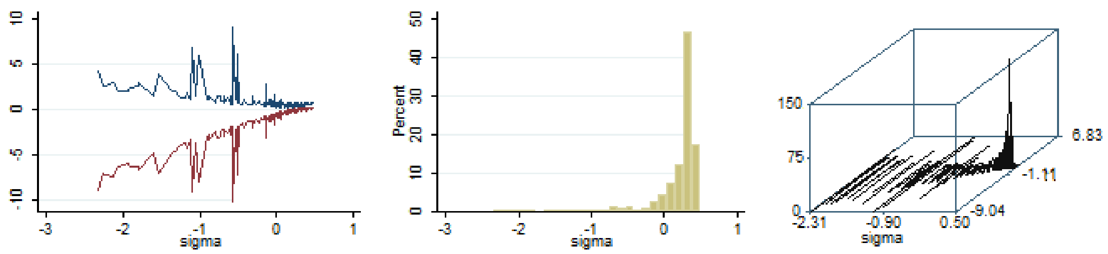

FIGURE 5.6: Translog estimated L-M Hicks elasticity graphical analysis

5.7 CES estimation

In this section we follow the recommendation obtained from the Wald test and the graphical analysis and estimate a nested CES. Indeed, in the fourth chapter we showed how direct non-linear estimation of the CES should be preferred in order to obtained the less bias results.

With a nested CES function three estimation methods have been used so far: the first is based on a non-linear estimation method (Kemfert, 1998, Koesler and Schymura, 2015), the second on the linearisation of the nested CES (Hoff, 2014) and the third on the estimation of the FOCs derived from a stepwise optimization procedure where a cost function based on the first the inner CES and then the one based complete nested CES are minimised (Chang, 1994, Prywes, 1986, van der Werf, 2008, Baccianti, 2013).

We use a direct estimation method and estimate the nested CES with a Maximum Likelihood estimator. We are aware that this is not the most efficient estimator given the econometric issues underlined by the diagnostic tests; however the obtained coefficients will be unbiased and consistent. The nested CES can be expressed with the following notation:[12]

$$lnQ_{it} = ln\lambda + \gamma t + \frac{\sigma}{\sigma - 1} ln\Big(\delta X_{it}^{\frac{\sigma-1}{\sigma}} + \delta_z L_{it}^{\frac{\sigma-1}{\sigma}} + (1 - \delta - \delta_z) M_{it}^{\frac{\sigma-1}{\sigma}}\Big) \tag{5.3}$$

with

$$X_{it} = ln\Big(\delta_x E_{it}^{\frac{\sigma_x-1}{\sigma_x}} + (1 - \delta_x) K_{it}^{\frac{\sigma_x-1}{\sigma_x}}\Big)^{\frac{\sigma_x}{\sigma_x-1}} \tag{5.4}$$

where $\lambda \in [0, +\infty)$ is the efficiency parameters, γ is a measure of technological progress, $\delta \in (0, 1)$, $\delta_x \in (0, 1)$ and $\delta_z \in (0, 1)$ are share parameters and σ and σ_x are substitution elasticities. We assume that the nested CES is characterised by constant returns to scale.

In Table 5.10 we report the results of the Maximum Likelihood estimation regression. We can see that all regressors are significant at a 5% level and that they lie in the ranges predicted by the economic theory. The elasticity of substitution between energy and capital is equal to 0.883. This is in line with our previous findings as it falls in the estimated prediction interval. The elasticity of substitution between the energy and capital composite input and the remaining inputs is equal to 0.468.

Parameters	**Coef.**	**Std. Err.**	**P**	**95% Conf.**	**Interval**
δ	0.476	0.025	0.000	0.427	0.526
δ_x	0.253	0.029	0.000	0.196	0.310
δ_z	0.156	0.018	0.000	0.121	0.191
σ	0.468	0.051	0.000	0.368	0.568
σ_x	0.883	0.440	0.045	0.021	1.745
λ	0.093	0.015	0.00	0.063	0.123
γ	0.003	0.001	0.00	0.004	0.002

TABLE 5.10: Maximum Likelihood estimation of the nested CES production function

5.8 Conclusions

In this chapter, we contribute to the applied econometric literature on the substitution relationships between inputs of production by estimating the elasticities of substitution between energy and other inputs. Our data are drawn from the EU-KLEMS database and

[12]In these equations we suppress the *it* subscript on each variable to slim down notation.

include 23 UK industrial sectors for the period 1970–2005. In line with the cited literature, we employ a Translog functional form to describe our production function. Furthermore, we compute three different types of elasticities: the Hicks, Allen and Morishima elasticities. Our results suggest that energy and capital are substitutes in production.

We also contribute to the CGE literature by providing both an indication of the appropriate nested structure and the relative constant elasticities for UK production. In the chapter, we check whether data support a nested CES representation of the production function. We use both empirical and graphical tests and we conclude that a nested structure of the form ((E,K),L,M) is the most appropriate to describe a CES production technology for the dataset under analysis. From the estimation of this nested CES, we obtain the constant elasticities of substitution which are equal to 0.88 and 0.47 for the inner and the outer nest respectively.

We conclude by briefly noting that thanks to the availability of long inputs and output time-series for a decent number of European countries, an interesting development of this research would concern testing separately for each industrial sector which ones is (are) the best nested structure(s) to describe the production function with a CES technology and for each of them estimate the relative constant elasticities. Indeed, the idea that the production technology is the same across all sectors is not realistic: the econometric literature shows how the distributions of Translog elasticities vary from industry to industry. The indication of the appropriate nested CES for each sector could be of particular interest for the CGE literature to better represent the production side of their economic models.

References

Allan, G., N. Hanley, P. McGregor, K. Swales, and K. Turner. 2007. "The impact of increased efficiency in the industrial use of energy: A computable general equilibrium analysis for the United Kingdom." *Energy Economics*, 29(4): 779–798.

Allen, R. G. 1934. "A comparison between different definitions of complementarity and competitive goods." *Econometrica*, 2(2): 168–175.

Allen, R. G. 1938. *Mathematical Analysis for Economists*. London: McMillan.

Anderson, R. G., and J. G. Thursby. 1986. "Confidence intervals for elasticity estimators in translog models." *The Review of Economics and Statistics*, 68(4): 647–656.

Apostolakis, B. E. 1990. "Energy-capital substitutability/complementarity." *Energy Economics*, 12(1): 48–58.

Arnberg, S., and T. B. Bjorner. 2007. "Substitution between energy, capital and labour within industrial companies: A micro panel data analysis." *Resource and Energy Economics*, 29(2): 122–136.

Arrow, K. J., H. B. Cheney, B. S. Minhas, and R. M. Solow. 1961. "Capital-labor substitution and economic efficiency." *The Review of Economics and Statistics*, 43(3): 225–250.

Baccianti, C. 2013. "Estimation of sectoral elasticities of substitution along the international technology frontier." ZEW Discussion Papers 13-092, ZEW - Zentrum für Europäische Wirtschaftsforschung/ Center for European Economic Research.

Baum, C. F. 2000a. "XTTEST2: Stata module to perform Breusch-Pagan LM test for cross-sectional correlation in fixed effects model." Statistical Software Components, Boston College Department of Economics, December.

Baum, C. F. 2000b. "XTTEST3: Stata module to compute Modified Wald statistic for groupwise heteroskedasticity." Statistical Software Components, Boston College Department of Economics, October.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2002. "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation." Statistical Software Components, Boston College Department of Economics, April.

Berndt, E. R., and L. R. Christensen. 1973a. "The internal structure of functional relationships: Separability, substitution, and aggregation." *The Review of Economic Studies*, 40(3): 403–410.

Berndt, E. R., and L. R. Christensen. 1973b. "The translog function and the substitution of equipment, structures and labor in U.S. manufacturing 1929-68." *Journal of Econometrics*, 1(1): 81–114.

Berndt, E. R., and D. Wood. 1975. "Technology, prices and the derived demand for energy." *The Review of Economics and Statistics*, 57(3): 259–268.

Berndt, E. R., and D. Wood. 1979. "Engineering and econometric interpretations of energy-capital complementarity." *The American Economic Review*, 69(3): 342–354.

Binswanger, H. P. 1974. "The measurement of technical change biases with many factors of production." *The American Economic Review*, 64(6): 964–976.

Blackorby, C., D. Primont, and R. Russel. 1977. "On testing separability restrictions with flexible functional forms." *Journal of Econometrics*, 5(2): 195–209.

Blackorby, C., and R. Russell. 1989. "Will the real elasticity of substitution please stand up? (a comparison of the Allen/Uzawa and Morishima elasticities)." *The American Economic Review*, 79(4): 882–888.

Burgess, D. F. 1975. "Duality theory and pitfalls in the specification of technologies." *Journal of Econometrics*, 3(2): 105–121.

Burki, A., and M. U. H. Khan. 2004. "Effects of allocative inefficiency on resource allocation and energy substitution in Pakistan's manufacturing." *Energy Economics*, 26(3): 371–388.

Chang, K. 1994. "Capital-energy substitution and the multi-level CES production function." *Energy Economics*, 16(1): 22–26.

Choi, I. 2001. "Unit root tests for panel data." *Journal of International Money and Finance*, 20(2): 249–272.

Christensen, L. R., and E. R. Berndt. 1973. "The internal structure of functional relationships: Substitution, and aggregation." *The Review of Economic Studies*, 40(3): 403–410.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1973. "Trascendental logarithmic production frontiers." *The Review of Economics and Statistics*, 55(1): 28–45.

Christev, A., and A. M. Featherstone. 2009. "A note on Allen–Uzawa partial elasticities of substitution: The case of the translog cost function." *Applied Economics Letters*, 16(11): 1165–1169.

Christopoulos, D. K. 2000. "The demand for energy in Greek manufacturing." *Energy Economics*, 22(5): 569–586.

Chung, J. W. 1987. "On the estimation of factor substitution in the translog model." *The Review of Economics and Statistics*, 69(3): 409–417.

Danny, M., J. May, and C. Pinto. 1978. "The demand for energy in Canadian manufacturing: Prologue to an energy policy." *The Canadian Journal of Economics*, 11(2): 300–313.

Dargay, J. M. 1983. "The demand for energy industries manufacturing." *The Review of Economics and Statistics*, 85(1): 37–51.

Davidson, R., and J. G. MacKinnon. 1981. "Several tests for model specification in the presence of alternative hypotheses." *Econometrica*, 49(3): 781–793.

Davidson, R., and J. G. MacKinnon. 1993. *Estimation and inference in econometrics*. Oxford: Oxford University Press.

Denny, M., and M. Fuss. 1977. "The use of approximation analysis to test for separability and the existence of consistent aggregates." *The American Economic Review*, 67(3): 404–418.

Despotakis, K. A., and A. C. Fisher. 1988. "Energy in a regional economy: A computable general equilibrium model for California." *Journal of Environmental Economics and Management*, 15(3): 313–330.

Diewert, W. E. 1971. "An application of the Shephard duality theorem: A generalized Leontief production function." *Journal of Political Economy*, 79(3): 481–507.

Diewert, W. E. 1973. "Separability and a generalization of the Cobb-Douglas cost, production and indirect utility functions."Technical Report 86, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Dissou, Y., L. Karnizova, and Q. Sun. 2015. "Industry-level econometric estimates of energy-capital-labour substitution with a nested CES production function." *Atlantic Economic Journal*, 43(1): 107–121.

Driscoll, J., and A. Kraay. 1998. "Consistent covariance matrix estimation with spatially dependent panel data." *Review of Economics and Statistics*, 80: 549–560.

Field, B. C., and C. Grebenstein. 1980. "Capital-energy substitution in U.S. manufacturing." *The Review of Economics and Statistics*, 62(2): 207–212.

Frondel, M. 2011. "Modelling energy and non-energy substitution: A brief survey of elasticities." *Energy Policy*, 39(8): 4601–4604.

Fuss, M. A. 1977. "The demand for energy in Canadian manufacturing." *Journal of Econometrics*, 5(1): 89–116.

Garofalo, G. A., and D. Malhotra. 1988. "Aggregation of capital and its substitution with energy." *Eastern Economic Journal*, 14(3): 251–262.

Greene, W. H. 2008. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Gregory, A. W., and M. R. Veall. 1985. "Formulating Wald tests of nonlinear restrictions." *Econometrica*, 53(6): 1465–68.

Griffin, J. M., and P. R. Gregory. 1976. "Intercountry translog model energy substitution responses." *The American Economic Review*, 66(5): 845–857.

Ha, S. J., I. Lange, P. Lecca, and K. Turner. 2012. "Econometric estimation of nested production functions and testing in a computable general equilibrium analysis of economy-wide rebound effects." Stirling Economics Discussion Papers 2012-08, University of Stirling, Division of Economics.

Hall, R. E., and D. W. Jorgenson. 1967. "Tax policy and investment behaviour." *The American Economic Review*, 57(3): 391–414.

Haller, S., and M. Hyland. 2014. "Capital-energy substitution: Evidence from a panel of Irish manufacturing firms." *Energy Economics*, 45(C): 501–510.

Hazilla, M., and R. J. Kopp. 1986. "Testing for separable functional structure using temporary equilibrium models." *Journal of Econometrics*, 33(1-2): 119–141.

Henningsen, A., and G. Henningsen. 2012. "On estimation of the CES production function - revisited." *Economics Letters*, 115(1): 67–69.

Hertel, T. W., and T. D. Mount. 1985. "The pricing of natural resources in a regional economy." *Land Economics*, 61(3): 229–243.

Hesse, D. M., and H. Tarkka. 1986. "The demand for capital, labor and energy in European industry manufacturing before and after the oil price shocks." *The Scandinavian Journal of Economics*, 88(3): 529–546.

Hicks, J. R. 1932. *Theory of Wages*. London: McMillan.

Hisnanick, J. J., and B. L. Kyer. 1995. "Assessing a disaggregated energy input using confidence intervals around translog elasticity estimates." *Energy Economics*, 17(2): 125–132.

Hoechle, D. 2006. "XTSCC: Stata module to calculate robust standard errors for panels with cross-sectional dependence." Statistical Software Components, Boston College Department of Economics, November.

Hoechle, D. 2007a. "Robust standard errors for panel regressions with cross–sectional dependence." *Stata Journal*, 7(3): 281–32.

Hoechle, D. 2007b. "Robust standard errors for panel regressions with cross-sectional dependence." *Stata Journal*, 7(3): 281–312.

Hoff, A. 2014. "The linear approximation of the CES function with n input variables." *Marine Resource Economics*, 19(3): 295–306.

Hoyos, R. E. D., and V. Sarafidis. 2006b. "XTCSD: Stata module to test for cross-sectional dependence in panel data models." Statistical Software Components, Boston College Department of Economics, June.

Hoyos, R. E. D., and V. Sarafidis. 2006a. "Testing for cross-sectional dependence in panel-data models." *Stata Journal*, 6(4): 482–496.

Hudson, E. A., and D. W. Jorgenson. 1974. "U.S. energy policy and economic growth 1975-2000." *The Bell Journal of Economics and Management Science*, 5(2): 461–514.

Hulten, C. R. 1990. "The measurement of capital." In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*. Eds. by E. R. Berndt, and J. E. Triplett: National Bureau of Economic Research Studies in Income and Wealth.

Ilmakunnas, P., and H. Torma. 1989. "Structural change in factor substitution in Finnish manufacturing." *The Scandinavian Journal of Economics*, 91(4): 705–721.

Iqbal, M. 1986. "Substitution of labour, capital and energy in the manufacturing sector of Pakistan." *Empirical Economics*, 11(2): 81–95.

Jorgenson, D. W., and Z. Griliches. 1967. "The explanation of productivity change." *The Review of Economic Studies*, 34(3): 249–283.

Kemfert, C. 1998. "Estimated substitution elasticities of a nested CES production function approach for Germany." *Energy Economics*, 20(3): 249–264.

Khiabani, N., and K. Hasani. 2010. "Technical and allocative inefficiencies and factor elasticities of substitution: An analysis of energy waste in Iran's manufacturing." *Energy Economics*, 32(5): 1182–1190.

Kim, J., and E. Heo. 2013. "Asymmetric substitutability between energy and capital: Evidence from the manufacturing sectors in 10 OECD countries." *Energy Economics*, 40(C): 81–89.

Kleibergen, F., and M. E. Schaffer. 2007. "RANKTEST: Stata module to test the rank of a matrix using the Kleibergen-Paap rk statistic." Statistical Software Components, Boston College Department of Economics, August.

Kmenta, J. 1967. "On estimation of the CES production function." *International Economic Review*, 8(2): 180–189.

Koesler, S., and M. Schymura. 2015. "Substitution elasticities in a constant elasticity of substitution framework: Empirical estimates using Nonlinear Least Squares." *Economic Systems Research*, 27(1): 101–121.

Koetse, M. J., H. L. de Groot, and R. J. Florax. 2008. "Capital-energy substitution and shifts in factor demand: A meta-analysis." *Energy Economics*, 30(5): 2236–2251.

Lafontaine, F., and K. J. White. 1986. "Obtaining any wald statistic you want." *Economics Letters*, 21(1): 35–40.

Lecca, P., K. Swales, and K. Turner. 2011. "An investigation of issues relating to where energy should enter the production function." *Economic Modelling*, 28(6): 2832–2841.

Li, P., and A. Rose. 1995. "Global warming policy and the Pennsylvania economy: A computable general equilibrium analysis." *Economic Systems Research*, 7(2): 151–171.

Maddala, G. S., and S. Wu. 1999. "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and Statistics*, 61(S1): 631–652.

Magnus, J. R. 1979. "Substitution between energy and non-energy inputs in the Netherlands 1950-1976." *International Economic Review*, 20(2): 465–484.

Mander, A. 2005. "SURFACE: Stata module to draw a 3D wireform surface plot." Statistical Software Components, Boston College Department of Economics, January.

Medina, J., and J. Vega-Cervera. 2001. "Energy and the non-energy inputs substitution: Evidence for Italy, Portugal and Spain." *Applied Energy*, 68(2): 203–214.

Moghimzadeh, M., and K. O. Kymn. 1986. "Cost shares, own, and cross-price elasticities in U.S. manufacturing with disaggregated energy inputs." *The Energy Journal*, 7(4): 65–80.

Morishima, M. 1967. "A few suggestions on the theory of elasticity." *Keizai Hyoron (Economic Review)*, 16: 144–150.

Nguyen, S., and M. Streitwieser. 1999. "Factor substitution in U.S. manufacturing: Does plant size matter?" *Small Business Economics*, 12(1): 41–57.

Norsworthy, J., and D. H. Malmquist. 1983. "Input measurement and productivity growth in Japanese and U.S. manufacturing." *The American Economic Review*, 73(5): 947–967.

Okagawa, A., and K. Ban. 2008. "Estimation of substitution elasticities for CGE models." Discussion Papers in Economics and Business 08-16, Osaka University, Graduate School of Economics and Osaka School of International Public Policy (OSIPP).

Ozatalay, S., S. Grubaugh, and T. Veach Long II. 1979. "Energy substitution and national energy policy." *The American Economic Review*, 69(2): 369–371.

Pedroni, P. 2000. "Fully modified OLS for heterogeneous cointegrated panels." Department of Economics Working Papers 2000-03, Department of Economics, Williams College.

Perroni, C., and T. F. Rutherford. 1995. "Regular flexibility of nested CES functions." *European Economic Review*, 39(2): 335–343.

Pindyck, R. S. 1979. "Interfuel substitution and the industrial demand for energy: An international comparison." *The Review of Economics and Statistics*, 61(2): 169–179.

Pindyck, R. S., and J. J. Rotemberg. 1983. "Dynamic factor demands and the effects of energy price shocks." *The American Economic Review*, 73(5): 1066–1079.

Prywes, M. 1986. "A nested CES approach to capital-energy substitution." *Energy Economics*, 8(1): 22–28.

Robinson, J. 1933. *The Economics of Imperfect Competition*. London: McMillan.

Roy, J., A. H. Sanstad, J. a. Sathaye, and R. Khaddaria. 2006. "Substitution and price elasticity estimates using inter-country pooled data in a translog cost model." *Energy Economics*, 28(5-6): 706–719.

Sato, K. 1967. "A two-level constant elasticity of substitution production function." *The Review of Economics Studies*, 34(2): 201–218.

Saunders, H. D. 2000. "Does predicted rebound depend upon distinguishing between energy and energy services?" *Energy Policy*, 28: 439–449.

Schaffer, M. E. 2005. "XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models." Statistical Software Components, Boston College Department of Economics, November.

Schaffer, M. E., and S. Stillman. 2006. "XTOVERID: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2, xthtaylor." Statistical Software Components, Boston College Department of Economics, October.

StataCorp. 2013. "Stata Statistical Software: Release 13." College Station, TX: StataCorp LP.

Strotz, R. H. 1959. "The utility tree: A correction and further appraisal." *Econometrica*, 27(3): 482–488.

Thompson, P., and T. G. Taylor. 1995. "The capital-energy substitutability debate: A new look." *The Review of Economics and Statistics*, 77(3): 565–569.

Thursby, J. G., and C. A. K. Lovell. 1978. "An investigation of the Kmenta approximation to the CES function." *International Economic Review*, 19(2): 363–377.

Turner, K. 2009. "Negative rebound and disinvestment effects in response to an improvement in energy efficiency in the UK economy." *Energy Economics*, 31(5): 648–666.

Turnovsky, M., M. Folie, and A. Ulph. 1982. "Factor substitutability in Australian manufacturing with emphasis on energy inputs." *Economic Record*, 58(1): 61–72.

Uzawa, H. 1962. "Production functions with constant elasticities of substitution." *The Review of Economics Studies*, 29(4): 291–299.

Voung, Q. H. 1989. "Likelihood Ratio tests for model selection and non-nested hypotheses." *Econometric Reviews*, 57(2): 307–333.

van der Werf, E. 2008. "Production functions for climate policy modeling: An empirical analysis." *Energy Economics*, 30(6): 2964–2979.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Zellner, A. 1962. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association*, 57(298): 348–368.

Zha, D., and N. Ding. 2014. "Elasticities of substitution between energy and non-energy inputs in China power sector." *Economic Modelling*, 38(C): 564–571.

Zha, D., and D. Zhou. 2014. "The elasticity of substitution and the way of nesting CES production function with emphasis on energy input." *Applied Energy*, 130(C): 793–798.